Evaluating Serious Slow Game Jams as a Mechanism for Co-Designing Serious Games to Improve Understanding of Cybersecurity

SHENANDO STALS, Heriot-Watt University, Edinburgh, United Kingdom of Great Britain and Northern Ireland
LYNNE BAILLIE, Heriot-Watt University, Edinburgh, United Kingdom of Great Britain and Northern Ireland
JAMIE IONA FERGUSON, School of Simulation and Visualisation, The Glasgow School of Art, Glasgow, United Kingdom of Great Britain and Northern Ireland
DAISY ABBOTT, School of Simulation and Visualisation, The Glasgow School of Art, Glasgow, United Kingdom of Great Britain and Northern Ireland
MANUEL MAAREK, Heriot-Watt University, Edinburgh, United Kingdom of Great Britain and Northern Ireland
RYAN SHAH, Heriot-Watt University, Edinburgh, United Kingdom of Great Britain and Northern Ireland
SANDY LOUCHART, School of Simulation and Visualisation, The Glasgow School of Art, Glasgow, United Kingdom of Great Britain and Northern Ireland

We present a first evaluation of a Serious Slow Game Jam (SSGJ) methodology as a mechanism for co-designing serious games in the application domain of cybersecurity, to evaluate how the SSGJ contributed to improving the understanding of cybersecurity. To this end, we engaged 13 participants with no experience in cybersecurity, in a multidisciplinary SSGJ involving domain-specific, pedagogical, and game design knowledge, and encouraged engagement in-between scheduled days of the SSGJ. Findings show improved confidence of participants in their knowledge of cybersecurity (from 12.5% to 62.5%) after undertaking the SSGJ, with free-text answers specifically indicating an improved understanding in terms of vulnerabilities, attacks, and defences for three quarters of the participants. Also, confidence in knowledge of game design improved (12.5% to 75%), and the SSGJ successfully engaged participants in-between scheduled days. Finally, a serious game is presented that was co-designed with participants during our SSGJ, and produced as an output of the SSGJ methodology.

CCS Concepts: • **Human-centered computing → HCI design and evaluation methods**; • **Security and privacy →** *Usability in security and privacy*; *Human and societal aspects of security and privacy*.

Additional Key Words and Phrases: Serious Slow Game Jam, Serious Games, Evaluation, Workload, Motivation, Engagement, Cybersecurity, Secure Coding, Secure Code Citizens

Authors' Contact Information: Shenando Stals, Heriot-Watt University, Edinburgh, United Kingdom of Great Britain and Northern Ireland; e-mail: s.stals@hw.ac.uk; Lynne Baillie, Heriot-Watt University, Edinburgh, United Kingdom of Great Britain and Northern Ireland; e-mail: l.baillie@hw.ac.uk; Jamie Iona Ferguson, School of Simulation and Visualisation, The Glasgow School of Art, Glasgow, United Kingdom of Great Britain and Northern Ireland; e-mail: J.Ferguson@gsa.ac.uk; Daisy Abbott, School of Simulation and Visualisation, The Glasgow School of Art, Glasgow, United Kingdom of Great Britain and Northern Ireland; e-mail: d.abbott@gsa.ac.uk; Manuel Maarek, Heriot-Watt University, Edinburgh, United Kingdom of Great Britain and Northern Ireland; e-mail: m.maarek@hw.ac.uk; Ryan Shah, Heriot-Watt University, Edinburgh, United Kingdom of Great Britain and Northern Ireland; e-mail: r.shah@hw.ac.uk; Sandy Louchart, School of Simulation and Visualisation, The Glasgow School of Art, Glasgow, United Kingdom of Great Britain and Northern Ireland; e-mail: s.louchart@gsa.ac.uk.

ACM 2832-5516/2025/1-ART
https://doi.org/10.1145/3709745

1 Introduction

With the democratisation of software development and deployment, the issues of code security and safety are widening. At the centre of this democratisation are the new code-citizens who are code-literate and able to build and run their own software code, by coding it themselves or by using code snippets from others [31]. They may have had no formal software engineering training and are often outside of the software industry which normally teaches, instils, and instructs best practices via house standards. Their understanding of the security implications of their coding is of fundamental importance to the security of software systems [31]. Research has shown that of the 1.3 million Android applications that contained security-related code snippets from Stack Overflow, 97.9% contained at least one insecure code snippet [24, 44]. We propose to put new code-citizens at the heart of secure code development, by using a Serious Slow Game Jam (SSGJ) methodology to engage them in the co-design of serious games. The SSGJ methodology implements an inclusive, collaborative, and creative framework for multidisciplinary teams. Compared to traditional, fast-paced game jams, it reduces the time pressure and inserts space for reflection on collaboration and play, the development of common-ground knowledge, and the exchange of knowledge. The SSGJ is spread across multiple weeks, allowing for reflection and refinement between scheduled days of the SSGJ. It provides support and mentorship from game design experts and cybersecurity experts throughout the SSGJ. At the same time, it retains the co-creation and supportive ethos of the traditional, fast-paced game jam. The SSGJ toolkit, comprised of a provocative game, Miro boards [66], Cybersecurity cards (for the application domain), Learning Mechanics cards, and Game Mechanics cards, is used to facilitate the SSGJ.

1.1 Research Questions

The overall aim of the research presented in this paper is to conduct a first evaluation of how our SSGJ methodology contributed to improving the understanding of cybersecurity. It investigates how different aspects of the Serious Slow Game Jam being evaluated may have contributed to this goal. To investigate this, we were guided by the following research questions:

- RQ1: How has the SSGJ affected participants' understanding of cybersecurity?
- RQ2: How can the cards for the application domain (in our case Cybersecurity cards), Learning Mechanics cards, and Game Mechanics cards that are part of the SSGJ toolkit, assist in serious game design?
- RQ3: What are the workload and motivation levels of participants during the SSGJ?
- RQ4: How has the "slow" format of the SSGJ affected participant engagement?

The remainder of this paper is organised as follows. Section 2 provides background knowledge regarding game jams, serious game design, the rationale for the proposed SSGJ methodology, and the potential evaluation methods for different aspects of the SSGJ. Section 3 presents the Serious Slow Game Jam methodology and the procedure for evaluating a Serious Slow Game Jam we conducted in the application domain of cybersecurity. The results of this evaluation are presented in Section 4. Section 5 discusses one of the serious games that was further developed by serious game design experts, from the co-designed (digital and non-digital) prototypes delivered through the SSGJ methodology. The findings are discussed in Section 6, and the conclusions and future work are presented in Section 7.

2 Background

2.1 Game Jams and Serious Games

Game jams have received attention from different research fields and perspectives [26]. Compared to game development in the entertainment industry, most game jams are typically aimed at creating rapid prototypes of entertainment games. According to Kultima [48], a game jam is an accelerated opportunistic game creation event where a game is created in a relatively short timeframe, during which design constraints are imposed and explored to inspire creativity, and the resulting outputs are shared publicly. Due to their purposefully short duration with an average range of 24-48 hours [3, 48], game jams provide a framework best suited for a rapid prototyping approach [3, 68]. Research has focused on various aspects of game jams, including the qualities of games produced during game jams [26], the development practices that participants employ [70, 98], and the educational value of game jam participation in terms of the development of soft and hard skills [29, 63, 75]. The impact of game jams as a cultural practice on the game developer community has also been investigated [56, 75], as well as the use of game jams as production and training tools in formal educational frameworks [25]. Game jams can have different effects on learning. They have been shown to enhance game development skills, STEM skills, and personal and interpersonal skills [63]. They also have been shown to enhance teamwork, communication, and project-management skill sets [29], provide an (online or co-located) environment that encourages social, participatory or situational learning [29], as well as self-directed learning [22]. They also enable formal and informal learning through play and experimentation [33, 62, 82], and can provide more abstract learning experiences that increase motivation and self-efficacy [9].

2.1.1 Serious Games.
Serious games have been defined as "games that do not have entertainment, enjoyment, and fun as their primary purpose" [64]. Most definitions of the concept are vague, and what their primary purpose is can differ. However, it is argued that the addition of pedagogy, that is, activities that educate or instruct and thereby imparting knowledge and skill, is what makes a game serious [32, 99]. Serious games can allow learners to experience situations that are difficult to experience in the real world for reasons of safety, cost, time etc., and have an impact on the player's development of different skills [32]. To this end, serious game design is multidisciplinary as it involves the synthesis of domain-specific knowledge, pedagogical knowledge, and game design knowledge, as reflected by the Triadic Game Design (TGD) methodology [40].

Game-based approaches have been applied for training, motivation and education in the domains of software engineering and cybersecurity [19, 59, 72]. One form of intervention is to simply 'gamify' existing tools or development methods, for example, using a Catch-The-Flag contest to train players to build, break, and fix software [71]. Investigations of gamification for software engineering [19, 72] outline the preliminary research and challenges, however it is advised to move beyond simple gamification approaches such as points and rewards [47, 72]. Therefore, bespoke serious games are considered to be promising opportunities for learning and training, and a range of games have been developed for both software engineering [59] and specifically for cybersecurity. Examples include competitive games where players attempt to find uncaught mutants or to write mutant-catching tests [83], a tabletop game on the drivers and biases within security decision-making [30], and a co-designed tower defence game focused on developer-centred security [4, 57, 58].

2.1.2 Serious Game Jams.
Serious game jams have been recommended and implemented as a tool for serious game design and research [18, 79]. Besides industry and independent/indie organized game jams, academic organized game jams are one of the three categories of game jams identified in the literature [33]. Academic game jams characteristically bring together academic researchers with the aim to produce research outputs. Compared to game jams focused on the creation of entertainment games, they replace abstract and arbitrary design themes with context-based research based on an application domain (e.g. health or cybersecurity). However, as has been previously noted, serious game jams have different needs in terms of their processes, personnel, and outcomes [3] and benefit from time and space for reflection and verification in addition to entertainment jams' key characteristics of the rapid prototyping approach (e.g. creativity through constraint, risk-taking, lower time/investment costs, and the selection of the best ideas for further development). Also, academic game jams are typically run by academic institutions or as part of academic conferences or workshops, which often limits the accessibility for and diversity

of participants [33, 49]. For example, to participate in Quantum Game Jams, participants had to be either scientists who are experts in the application domain, or experienced game designers or game developers [50].

In addition, in serious game design, the gameplay design needs to be understood with regards to the application domain and pedagogical objectives, conceptualised in Triadic Game Design [41] as Reality (i.e. the application domain), Meaning (i.e. pedagogical value), and Play (i.e. gameplay). Therefore, serious games design places much more emphasis on mapping these three aspects for effective learning outcomes. However, the conventional entertainment-oriented game jam format is not best suited for the needs of serious game design and serious game design research [6]. Their generally fast and intensive pace can make them inaccessible [26, 35, 49, 52] and leaves little room for participants to refine or reflect on their work [3]. This need for refinement and reflection was also identified by Hecker and Blow [43] who argued that game jams are shallow and horizontal by design. They introduced more vertical Depth Jams, in which developers focused on improving and polishing one particular feature or mechanic in an already existing game, in an attempt to bridge the gap between game jam (prototypes of) games and fully developed games [43, 52]. To address the need for deep understanding and mapping of game mechanics to learning outcomes and the explicit representation of pedagogical experts in serious game jams (in addition to subject and game experts), we have proposed the design of a new Serious Slow Game Jam method [3]. We define the SSGJ method by the key concepts of collaboration, improvisation, rapid creativity common to rapid prototyping approaches and design sprints. The SSGJ method incorporates these characteristics alongside the time-limited, stand-alone nature of a jam, however, instead of being intensely concentrated, the limited time ismeaningfully distributed over a longer period [3].

2.1.3 Game Jam Design Parameters.

Game jams can have diverse formats depending on their aims and contexts. However, an analysis of the literature on game jams identified several shared design parameters [21, 28, 51, 68]. These are the theme of the game jam, time constraints (i.e. typically ranging from 8-72 hours), the location (i.e. physical, online, or hybrid), participation and team requirements (i.e. prior experience and skills), technology use (i.e. technology-agnostic or dedicated platforms for games production), participant support (i.e. keynote talks, workshops, presentations, mentoring), and deliverables (e.g. game prototypes, supporting multimedia, and documentation) [3, 21, 28, 51, 68].

Past serious game jams have provided valuable recommendations regarding these design parameters, as well as insights into challenges, practicalities, and good practices regarding serious game jam design [1, 6, 76, 79]. Like academic-organized game jams, the theme for serious game jams does not just provide inspiration for game design, but also defines a research context and application domain. To achieve the research outputs requires accurate and appropriate content creation and thus expertise of the application domain. In addition, for the design of effective serious games, this requires significant expertise in pedagogy besides skills and expertise in game design and game development [3]. Recommendations from existing serious game jams advise to address these significant additional needs for participants via the Time and Support parameters [6, 76]. It is recommended to divide the game jam into phases with breaks in between [6], and explicitly provide (synchronous and/or asynchronous) educational content over a longer period of time [76]. It is also recommended to integrate domain experts (and other stakeholders) into the game jam as mentors or participants [1, 6, 76], and integrate lecture-like content at key points during the game jam. This is to ensure the alignment of goals of the activities, provide a structured journey for participants, and to ensure the validity of the content in the serious games produced [1, 76]. How this has informed the design of the Serious Slow Game Jam method [3] will be discussed in Section 3.1

2.2 Evaluating Game Jams

Evaluation of game jams has focused on various aspects of game jams, including the qualities of games produced during game jams [26], and the educational value of game jam participation in terms of the development of soft and hard skills [29, 63, 75]. More formal frameworks for evaluation of certain aspects of game jams have been suggested, such as the Triadic Game Design Evaluation framework for a balanced evaluation of serious games design [40]. However, game jam evaluations tend to focus on a specific aspect of the game jam [40], rely mostly on qualitative evaluation [27], and often lack rigour and depend on the evaluation of experts (sometimes based upon predetermined criteria) [40]. There is a gap in establishing a universal method with which to evaluate game jams against their intended outcomes [40, 80]. It has been suggested that, in comparison to traditional, short-format game jams, more extensive, in-depth game jams have potential for a more comprehensive evaluation [82].

2.2.1 Knowledge and Understanding.

Assessing learning outcomes from game jams is difficult [9, 29]. The learning experience is a very private experience [29], while devising tests for participants to assess their knowledge and understanding or having participants write reports after the game jam, add to the participant's workload and can take away from the playful learning and learning through experimentation that game jams provide [33, 62, 82]. Therefore, reports of learning are typically self-assessed by game jam participants [8, 9]. In order to determine how the SSGJ has affected participants' understanding of the application domain (in this case cybersecurity), pre-/post-tests, self-assessment, and peer-assessment can be used. Pre-/post-tests are one of the most used experimental designs in educational research to assess the effect of new teaching methods [10, 20]. This can be a one-group pre-/post-test design, or a pre-/post-test design where the results of the pre-/post-tests for a new teaching method, like learning by game design, are compared to the pre-/post-test results of a control group using an existing teaching method [78]. They have also been used in the context of serious games to assess learning outcomes after playing a serious game, with the pre-test providing a baseline to compare the test scores of the post-test against [10, 12, 42, 73]. It is also straightforward to implement both in paper or digital form, meaning it may be suitable for a SSGJ that takes a hybrid format. It provides a direct way to assess the learning outcomes of the SSGJ relating to levels of knowledge and understanding in cybersecurity [10, 12], and confidence in key cybersecurity skills [13, 31]. Besides obtaining knowledge, learning meta skills (e.g. socialising) and experiences of learning in game jams is also important, in particular from the perspective of motivation. These learning experiences in game jams can drive future learning and have a positive impact on self-efficacy [9, 63]. Confidence in key skills is an important aspect of the learning experience [13]. The Student Instrument for measuring Confidence in Key Skills (SICKS) [13] can be used for a quantitative measurement of confidence regarding cybersecurity and game design knowledge and skills. The SICKS measurement assesses confidence levels of students across six variables corresponding with key skills in education. For the evaluation of the SSGJ, those key skills could be replaced by key cybersecurity skills identified by Georgiou et al. [31]: code practices, resources, communication, and morality.

A potential limitation of pre-/post-tests is that unlike in-process assessment, this evaluation does not depend on all of the information that can be collected during and within the different phases and activities of the Serious Slow Game Jam [90], but additional methods in the form of self- and peer assessment of serious games can be implemented to this end [61]. Peer assessment has been shown to be more effective on a team level, as opposed to on an individual level [23], and in the context of serious game research can result in participants assessing their team skills more accurately [12]. For the assessment of participants' understanding of cybersecurity during the SSGJ, feedback could be provided to the participants which would contribute to their learning experience. This has been shown to be particularly important for participants unfamiliar with game design or the domain in which it is applied [14]. This feedback could be provided by the cybersecurity experts [94] and/or by other participants through peer-review of other participants' games [40]. Troiano et al. [94] advise that for serious games, the emphasis should be on evaluating the Reality construct of TGD (in particular reality representation and contextualization), and that this evaluation should be done by domain experts. However, Harteveld [40] stresses the importance of the player or learner being part of the evaluation. Advantages of peer assessment are that it helps participants understand their own work better, provides more immediate feedback, and improves attitudes towards the learning process [61, 87]. Therefore, peer assessment combined with feedback from experts would be the preferred evaluation method, which has been used in computer science education and in the software development process and for code reviews (e.g. [61]).

2.2.2 Workload.

A review of game jams over the past twenty years [52] revealed that traditional, fast-paced game jams aim to create a game in between one and three consecutive days. Although participants are typically people with experience in programming and game design, the schedule and strict deadlines result in a high workload [26, 35, 52]. To reinforce accessibility and inclusivity, the SSGJ model aims to be a 'no-crunch' working environment by having session durations that are non-exhausting, and re-evaluating time pressure based on serious game design needs [3]. To evaluate this, the workload of each of the activities during the SSGJ needs to be assessed. In game jams, the workload is typically evaluated qualitatively based on observations of the game jam organizers, or interviews with game jam participants [62]. However, the NASA-TLX may be a suitable option, as it is a subjective, multidimensional assessment that rates perceived workload [38, 39] quantitatively making it more suitable for comparison, and has been widely used in complex socio-technological domains like aviation and healthcare [16] and in the domain of gaming [53, 55, 77, 97]. Alternatives to measure workload are the Subjective Workload Assessment Technique (SWAT) [81] and Workload Profile (WP) [95]. However, these are not as well established as the NASA-TLX, are less robust, and are more difficult to comprehend by participants (in particular the WP) [84].

2.2.3 Motivation and Engagement.

Motivation plays an important role in participating in game jams [17]. Due to the "slow" aspect of the SSGJ, participants are asked to commit over a longer overall duration than a traditional game jam [52], while the contact hours remain the same. Therefore, understanding what motivates participants to participate and return (or drop out), is important as (intrinsic) motivation is more likely to enhance performance and persistence in an activity than extrinsic motivation [86]. Like the workload, this can be evaluated qualitatively using observations, questionnaires, and interviews [17, 62, 75]. This may also be assessed quantitatively using the Intrinsic Motivation Index (IMI) [60, 85] for each day of the SSGJ. The IMI is not domain-specific and has been used to this end in user-centred HCI design studies before (e.g. [96]). To evaluate the engagement in between SSGJ scheduled days, an evaluative questionnaire with open-ended questions at the end of the SSGJ may be a good option, as this is the most common tool used to evaluate game jams [9, 52].

3 Methodology

3.1 Serious Slow Game Jam Method

Based on the requirements for serious game design, recommendations from previous serious game jams, and the limitations of conventional entertainment-oriented game jam format identified in Section 2, we have proposed a Serious Slow Game Jam (SSGJ) methodology [3] which provides a multidisciplinary collaborative framework for serious game design, putting participants and experts at the centre of the design. It provides mentorship by application domain and game design experts to support participants, to support the value and validity of outputs, and to provide a structured, accessible, and educational experience. The SSGJ methodology differs from similar methods such as Quantum Game Jams (QGJ) [50] and Depth Jams [43]. QGJs address the need to introduce and present the application domain as well as the discipline of serious game design, but the participants are either experts in the application domain or in game design, and the QGJ format is particularly suited towards experienced game jammers. The 'meaning' aspect of Triadic Game Design is also not explicitly represented, whereas it is a crucial aspect of both process and mentorship in the SSGJ. In the SSGJ method the participants are not required to be experts in the application domain or (serious) game design, and might not have experience with game jams either [3]. Depth Jams [43] are spread over a longer period (i.e. 4 days) to allow time for reflection and refinement, but the participants are all game developers and focus is on refining (aspects of) an already existing game, whereas the SSGJ focuses on ways to rigorously define learning outcomes and then select and map appropriate game mechanics to deliver them. Although the SSGJ method is applied here to the application domain of cybersecurity, it is intended to be flexible and generic so that it can be used irrespective of application domain [3].

The SSGJ methodology falls under the wider category of 'applied game jams', which are defined as "game jams that explore a range of different topics, issues, and objectives through game development" [80]. For this reason, we have adopted the applied game jam framework [80] which involves investigation of four aspects: the problem space, the jam design, the jam delivery and its outcomes, and any follow-on opportunities.

3.1.1 Problem space.
In terms of the problem space, the aim was to enhance participants' understanding of the application domain, in our case cybersecurity, and reflect on the serious game designs [3].

3.1.2 Serious Slow Game Jam Design.
The game jam design is discussed in terms of the shared game jam design parameters [21, 28] identified in the literature and discussed in Section 2. The **Theme** is guided by domain expert mentors and structured educational materials. These are included in the SSGJ toolkit to facilitate the SSGJ (see the Support parameter) [3]. The cybersecurity theme for the specific SSGJ being evaluated here, is that of secure software development lifecycles.

For the **Time** design parameter, emphasis is on accessibility and inclusion, aiming for a non-crunch working environment with non-exhausting session durations for each day of the SSGJ [3, 49]. Based on lessons learned from previous serious game jams [6], the SSGJ is structured into three phases in the serious game creation lifecycle (i.e. design, development, and pre-release), consisting of two work days each, resulting in six days in total (see Table 1). The first phase consists of an introduction to the application domain and TGD. During this introduction phase presentations by domain experts, a deck of Cybersecurity, Learning Mechanics, and Game Mechanics cards (describing concepts within each topic, and are used throughout all days of the SSGJ), and a small provoking game are introduced to kick off discussions about the application domain. Phase two covers the design of the serious game loop and prototype design, and phase three covers the development of the serious game prototype and other deliverables. There is ample time in between each of these phase for reflection, feedback, and refinement (2 weeks), resulting in an overall duration for the SSGJ of 5 weeks [3].

For the **Location** parameter, due to the timing of this SSGJ which took place in spring 2022, the COVID-19 pandemic at the time, and the diversity of our target group, we wanted to see if we could run the SSGJ in a hybrid format, where participants may be in-person or synchronously online to prioritize inclusivity and accessibility [3].

For **Participation and Teams**, we followed recommendations form the literature [27] with the organizers of the SSGJ creating teams based on self-identified roles collected during participant registration. Where serious game research is an intended outcome, explicit inclusion of both domain experts and serious game designers as participant-mentors allows for the delivery of high-quality support materials (see [6, 25, 76]), guidance in framing the SSGJ theme [76], supporting and contextualizing domain related material, and validating its inclusion in the serious game [3]. Application domain experts in game design and cybersecurity rove between teams, to enhance contact and knowledge exchange between experts and participants [6, 76].

In addition to in-person communication and collaboration, **Technology** in the form of Discord [45] is used for online communication and Miro [66] with structured activity worksheets for online collaboration. Due to the wide and diverse skill set of the target audience, there are no limitations imposed for game platforms [3].

Regarding the **Support** of participants, the SSGJ methodology includes guided educational group activities, supported by physical and digital materials in the SSGJ toolkit [3]. The TGD method [41] is used to inform and guide participants with respect to serious game design. We propose this will result in strong learning outcomes for participants as well as serious game prototypes that have high rigour and domain validity [3]. The SSGJ toolkit to support the SSGJ includes presentations of domain experts, a provocative game [2], Miro boards [66],

Table 1. SSGJ structure showing time (in minutes), milestones, and evaluation activities.

Phase	Activity	Rationale	Procedure	Duration (minutes)
Prep	Participant information sheet	Inform participants	Online registration	5
	Informed consent form	Inform consent	Online registration	5
	Demographic Questionnaire	Participant profiles	Online registration	5
	Adjusted IMI	Motivation pre-SSGJ	Online registration	5
1	**Day 1**: Introduction to cybersecurity through provocative game and expert presentations. Introduction to TGD. TGD Reality session where cybersecurity issues the games will address are identified using the cybersecurity cards.			
	Pre-Test Questionnaire	Assess understanding.	Discord	10
	TLX: Day 1 (See Table 3)	Measure workload	Paper / Discord	4 x 5
	IMID: Day1	Measure motivation	Discord	5
	Day 2: TGD Meaning session: suitable learning mechanics are selected using the cards. TGD Play session: suitable game mechanics are selected using the cards and matched with learning mechanics.			
	TLX: Day 2 (See Table 3)	Measure workload	Paper / Discord	3 x 5
	IMI: Day 2	Measure motivation	Discord	5
2	**Day 3**: Introduction to game loops. The Serious Game Loop is designed and a paper prototype of the game is created.			
	TLX: Day 3 (See Table 3)	Measure workload	Paper / Discord	4 x 5
	IMI: Day 3	Measure motivation	Discord	5
	Day 4: Participants playtest their own paper prototype. Second round of paper prototyping.			
	TLX: Day 4 (See Table 3)	Measure workload	Paper / Discord	3 x 5
	Cybersecurity Cards Questionnaire	Evaluate cards	Discord	10
	Learning Cards Questionnaire	Evaluate cards	Discord	10
	Game Cards Questionnaire	Evaluate cards	Discord	10
	IMI for Day 4	Measure motivation	Discord	5
3	**Day 5**: Two development sessions to create the serious game.			
	TLX: Day 5 (See Table 3)	Measure workload	Paper / Discord	2 x 5
	IMI: Day 5	Measure motivation	Discord	5
	Day 6: Final development session. Self- and peer-assessment of serious game with expert feedback.			
	TLX: Day 6 (See Table 3)	Measure workload	Paper / Discord	4 x 5
	IMI: Day 6	Measure motivation	Discord	5
Post	Post-Test Questionnaire	Assess understanding	Discord	10
	SSGJ Experience Questionnaire	Evaluate SSGJ format.	Discord	30

and three decks of cards: Cybersecurity [89] (i.e. for the application domain) (Figure 1), Learning Mechanics (LM) (Figure 2), and Game Mechanics (GM) (Figure 3).

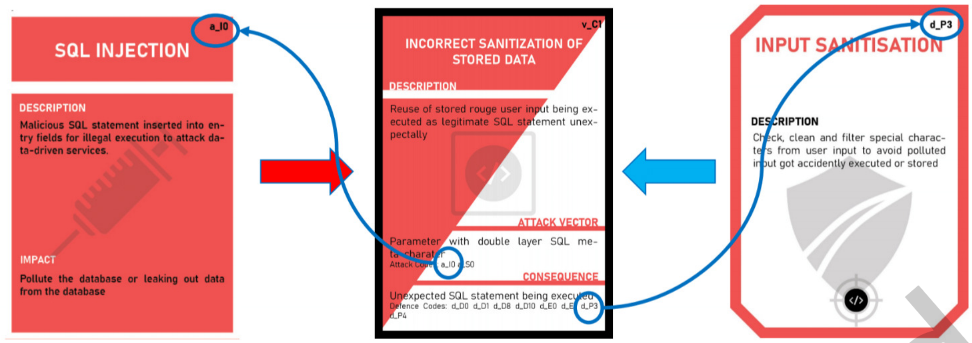

Fig. 1. Design of different types of Cybersecurity cards, with a detailed Attack card (left), Vulnerability card (middle), and Defense card (right) [89].

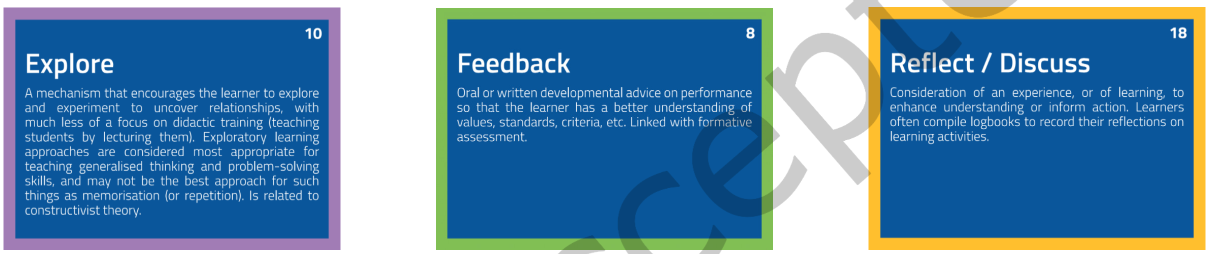

Fig. 2. Example cards of the Learning Mechanic deck based on the LM-GM framework by [54].

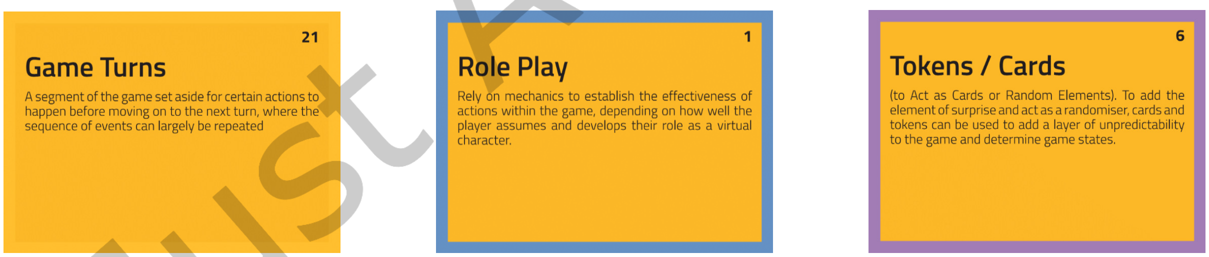

Fig. 3. Example cards of the Game Mechanic deck based on the LM-GM framework by [54].

The cybersecurity cards [89] have been developed based on CyBOK, a comprehensive body of knowledge to inform and underpin educational and professional training for the cybersecurity sector [69], while the LM and GM cards were created based on the LM-GM framework [54]. The Cybersecurity cards consist of general and more detailed Attack cards, Defence cards, and Vulnerability cards, which are related but not necessarily in a one-to-one relationship (Figure 1). The full decks of Cybersecurity cards, LM cards, and GM cards, which are part of the SSGJ toolkit, are freely available from our Secrious project website (https://secrious.github.io/#cards)[1].

[1]Please note that the Secrious project website contains the final version of the Cybersecurity cards, LM cards, and GM cards, as discussed in more detail in Section 6.2 and by [89].

Informed by recommendations from the literature [21], **Deliverables** for each team include a serious game prototype, and a Serious Game Design Document (SGDD) which lays out the serious game design according to a provided template, which all SSGJ activities feed into [3].

3.1.3 Serious Slow Game Jam Delivery and Outcomes.

For the SSGJ delivery and outcomes, the aim is to provide an educational experience for participants, which is engaging, and reduces the time pressure. Outcomes of each SSGJ include (digital or non-digital) serious game prototypes, SGDDs, an analysis of each serious game prototype, the participants' knowledge and understanding of the application domain (i.e. cybersecurity), game design/development, and their confidence levels in keys skills in those areas [3].

3.1.4 Follow-on Opportunities.

Through the integrated educational and analytical activities and the SGDD, the SSGJ methodology enables the subsequent additional analysis of SSGJ outputs by the researchers (e.g. pedagogical patters suited to problems in particular application domains). The overall SSGJ methodology outputs also provides a plethora of digital and physical resources: The design of the SSGJ methodology and the SSGJ toolkit (e.g. the Cybersecurity, Learning Mechanics, and Game Mechanics card decks) can be adapted and reused freely to support future SSGJs in different application domains [3].

The theoretical context and the concept, definition, and design parameters of the SSGJ methodology, as well as their trade-offs and limitations in comparison to other types of game jams, are presented in more detail in a previous publication [3]. The focus of this paper is on a first evaluation of the SSGJ methodology as a mechanism for co-designing serious games to improve understanding of cybersecurity, and the presentation of a serious game that was co-designed with participants during a SSGJ as an output of the SSGJ methodology. Table 1 shows and outline of the SSGJ schedule and the evaluation activities, which will be discussed in more detail in the next section.

3.2 Evaluation Procedure for the Serious Slow Game Jam Method

Our aim is to evaluate the Serious Slow Game Jam methodology as a mechanism for co-designing serious games to improve the understanding of cybersecurity. This evaluation is guided by the research questions outlined in Section 1.1, and informed by the parameters of the SSGJ design, and the discussion of different methods for evaluating game jams in Section 2.2.

3.2.1 Measuring Knowledge, Workload, Motivation and Engagement.

Participants' level of understanding of cybersecurity and game design and development and their confidence in key skills in those areas, was collected before and after the SSGJ using a one-group pre-/post-test questionnaires (see Section 2.2.1). It consisted of seven 7-point Likert questions, with the key skills in the SICKS questionnaire [91] being replaced by the key skills in cybersecurity [31] and game design/development (see Section 2.2.1). It was administered individually online using Microsoft Forms [65] at the start of Day 1 and at the end of Day 6 of the SSGJ. A one-group pre-/post-test design was chosen over a pre-/post-test design with a control group, due to the timing and COVID-19 restrictions of the SSGJ at the time, the small number of participants recruited, and offering of an alternative teaching method besides the SSGJ not being practically and logistically feasible [78]. Self-assessment and peer assessment at team level combined with feedback from experts (see Section 2.2.1) has been chosen for the qualitative assessment of participants' understanding of cybersecurity and serious game design and development during the SSGJ. Each team first assessed their own game using the three decks of cards (see Figures 1, 2, and 3) and discussed and explained their selection to at least one of the experts. The game was then played by another team, who also assessed it using the three decks of cards. The full decks of cards were provided in physical format and in digital format on the Miro board.

Workload was assessed using the NASA Task Load Index (NASA-TLX) [38, 39] (see Section 2.2.2). It was administered individually at the end of each activity, using pen/paper for in-person, and using Adobe Forms [5] for online participants.

Motivation to adhere and complete activities for each day of the SSGJ was measured using the Intrinsic Motivation Index (IMI) [60, 85] (see Section 2.2.3). The subscales of "effort" and "felt pressure" were removed, as these are already measured via each activity's NASA-TLX. This does not compromise the construct validity or internal consistency for the subscales [36]. The IMI was administered individually by sharing a Microsoft Forms link via the dedicated Discord channel at the end of each day of the SSGJ.

Finally, an open-ended questionnaire was used to encourage participants to reflect on the "slow" format of the SSGJ and to evaluate their engagement between scheduled days of the SSGJ (see Section 2.2.3). It was administered individually by sharing a Microsoft Forms link on the final day of the SSGJ.

3.2.2 Cybersecurity, Learning Mechanics and Game Mechanics Cards.

Inspired by the SICKS [13] (see Section 2.2.1), a questionnaire to evaluate how each deck of cards was used during the SSGJ has been created. Based on input from the cybersecurity experts who designed the Cybersecurity cards and the game design experts who designed the LM and GM cards for the SSGJ, these questionnaires enabled participants to individually self-assess to what degree, if any, each deck of the cards has contributed to: providing a knowledge base (including, cybersecurity/LM/GM concepts, scope, relationships between concepts, and terminology); Independent learning and self-efficacy; Inspiration for design of the serious game; Communication with experts and others; and providing a reference point for grounding cybersecurity issues/LM concepts/GM concepts during game design and development. Items related those aspects listed above were scored on a 7-point Likert scale (1=strongly disagree, 7=strongly agree). In addition, using checkboxes on the form, participants selected the activities of the SSGJ during which they used each deck of cards the most, the subset of types of cards within a deck they did not use, and which potential improvements to the design of the deck of cards should be made in their opinion. These three questionnaires (one for each deck of cards) were administered online by sharing Microsoft Forms link on Day 4 of the SSGJ via a dedicated Discord channel.

Table 1 shows the schedule for the SSGJ, and how the evaluation measures were integrated into the three phases across the six days of the SSGJ. Ethical approval was obtained from the researchers' university ethics board. Participants were asked before the start of the jam to register online, this included: participant information sheet, informed consent form, demographics questionnaire and adjusted IMI to assess motivation prior to the SSGJ.

4 Results

In this section, the results of the evaluation of the SSGJ are presented. The findings will be discussed in Section 6. The questionnaires used in the evaluation are made publicly available via a link to the Open Science Framework digital repository: https://doi.org/10.17605/OSF.IO/2MXDK [92]

4.1 Participants

Sixteen participants were recruited on campus (Table 2). Only the data of participants that actively participated in at least 4 out of 6 days of the SSGJ have been included, resulting in data of 13 participants in 3 teams being analyzed. Due to the timing of this SSGJ which took place in spring 2022 and the COVID-19 pandemic at the time, we wanted to see if we could run the SSGJ in a hybrid format, where participants may be in-person or synchronously online. Participants were mainly there in person, but five of the participants opted to partially take part in the SSGJ online (see Table 2). Activities for each day of the SSGJ lasted from 10:00-16:00 (GMT).

Table 2. Overview of attendance during SSGJ, with maxima for participant attendance highlighted in bold underlining.

Day	Participant (In-person)	Participant (Online)	Participant (Part day)	Expert (In-person)	Expert (Online)
1	10	3	0	7	1
2	11	2	2	7	1
3	9	**4**	3	4	2
4	8	2	**6**	7	1
5	8	1	**6**	7	1
6	**13**	0	3	7	1

Participants were aged between 22 and 35 years (mean 26.3 years), (3 female, 10 male). All participants were MSc conversion students in computer science (but not in cybersecurity), and came from an electrical or mechanical engineering (7), mathematics (3), computing (2), or biology background. One participant indicated having specific experience with secure coding, two indicated having experience with gaming, and one participant indicated having intermediate skills as a game developer in Java. None had participated in any type of game jam before. During the SSGJ, an expert in cybersecurity and in game design would always be present [3]. For this specific SSGJ, at most 7 experts were present in person and at least 1 expert was present online (Table 2), and there were in total 4 cybersecurity experts, 3 game design experts, and 2 HCI experts. They would periodically go round different teams to check if they had any questions or wanted to have discussion with an expert, but the serious games were co-designed among participants (i.e. experts were not part of a team).

4.2 Pre vs. Post Questionnaire

The 7-point Likert-scale data was categorized by the percentage of participants who reported they were confident (i.e. scored it 5/6/7 out of 7), neutral (i.e. scored it 4/7) or not confident (scored it 1/2/3 out of 7). These percentages were then compared pre- and post SSGJ [73, 91]. In addition, a Wilcoxon signed ranked test [7] was done to determine if there was a statistical difference between the pre- and post-test scores. The free-text answers were independently coded by two coders and grouped into themes, an analysis technique that was used in previous HCI research (e.g. [15, 67, 83]. Codes were analysed and verified by two postdoctoral researchers, one with expertise in HCI who had taken part in the SSGJ, and one with expertise in Cybersecurity who had not taken part in the SSGJ. Both are experienced in free-text coding qualitative data from questionnaires. Open discussion was used to systematically discuss and resolve the codes to reach consensus for the final coding [15].

4.2.1 Cybersecurity (Pre vs. Post SSGJ).

The Wilcoxon signed rank test [7] showed participants' confidence in their knowledge and understanding of cybersecurity in the post test scores had improved significantly compared to the pre test scores (Z=-2.041, p=0.041). Responses relating to *Code Practices* indicated confidence in current level of knowledge and understanding of cybersecurity shifted positively from 12.5% to 62.5%. Confidence with reviewing and updating existing code regarding cybersecurity dropped from 62.5% to 50%. Responses relating to *Resources* indicated increased confidence to ask for more money or funding to improve code security from 62.5% to 75%. Responses relating to *Communication* showed confidence in raising a security issue with their non-expert manager increased from 37.5% to 62.5%. Responses for *Morality* showed confidence to bring up a security issue that will knowingly have implications for the end user increased from 50% to 62.5%. Confidence to go against your manager when volume of output is prioritized over a security issue decreased from 50% to 37.5%. Other items, such as asking for an increase in staff to improve code security, showed little change.

4.2.2 Game Design and Development (Pre vs. Post SSGJ).

The Wilcoxon signed rank test also showed participants' confidence in their knowledge and understanding of game design and development in the post test scores had improved significantly compared to the pre test scores (Z=-2.112, p=.035). Regarding game design, confidence in current level of knowledge and understanding of game design (12.5% - 75%), and the ability to design a game (37.5% - 75%) shifted positively. Regarding game development, confidence in current level of knowledge of game development (12.5% - 62.5%), and in ability to implement a game (37.5% - 50%) shifted positively. Other items, such as teaching others about game development, showed little change.

4.2.3 Free-Text Responses.

The free-text responses provided some additional insights in the quantitative data presented above. Three quarters of participants self-reported the main thing they had learned about cybersecurity through the SSGJ was the different types of vulnerabilities, attacks, and defenses: *"I learned [..], some types of attacks, defenses, and vulnerabilities".* - (P7). Half of participants reported factors influencing cybersecurity (e.g. human factors), and almost a third mentioned the relationships between vulnerabilities, attacks, and defenses as well as terminology. There was an almost equal split between participant indicating they had learned a lot about cybersecurity during the SSGJ, and that they had not learned that much and wanted to learn more: *"[The SSGJ matched] 90% of my expectations [regarding learning about cybersecurity]. And personally I feel I have to learn a lot on cybersecurity. But the workshop has more than enough knowledge to get you started and understand cybersecurity."* - (P1). A third of participants felt the emphasis was too much on serious game design instead of learning about cybersecurity: *"Cybersecurity felt secondary to the concepts of serious games."* - (P2).

4.3 Workload

The raw NASA-TLX workload data was analyzed by taking the mean value of all responses for each activity and each workload subscale [38]. The average workload values for each of the subscales can be found in Table 3, namely mental demand (MD), physical demand (PD), temporal demand (TD), effort (EFF), frustration (FRU), and performance (PER). The scores were classified as Low (1-3), Medium (3-7), Bit High (7-11), High (11-17), and Very High (17-21) [38, 74].

The highest workload came from the three development sessions in Days 4 and 5, with mental demand (MD) ranging from 15.3-17.1 and effort (EFF) ranging from 14.7-15.8. The lowest mental demand (MD = 5.5) and effort (EFF = 6.8) came from the first icebreaking and team-formation activity during the introduction on Day 1. Looking at the various subscales across all activities, mental demand (MD) and effort (EFF) are classified as "high" throughout the SSGJ, but temporal demand (TD) is only classified as "high" (TD = 11.6-12.7) during the development activities in Phase 3 (Day 5 & 6), and during Serious Game Loop Design on Day 3 (11.8). Performance levels (PER), were highest (indicating participants felt they performed badly) during the activity to extract cybersecurity metaphors from the provoking game and during the three TGD activities (Reality, Meaning, and Play session).

4.4 Motivation

The IMI data was analyzed by averaging each subscale for each day of the SSGJ, and then refined further by looking at the percentage of participants who scored very highly (6/7 or 7/7) on each of the subscales (as per: [37, 46, 96]). Results in Table 4 shows that Interest/Enjoyment and Perceived Value/Usefulness are all positive (≥4.5/7) on average and stay positive throughout the SSGJ, indicating the SSGJ had successfully engaged participants on the topic of cybersecurity. Perceived Competence is average prior to the start of the SSGJ, but becomes positive

Table 3. Overview of the workload for each of the NASA-TLX subscales and the average per activity of the SSGJ, with the top-5 highest values in the "high" and "very high" classifications per subscale highlighted in bold underlining.

Day	Activity	MD	PD	TD	EFF	FRU	PER
Day 1	Icebreaking & Team formation	5.5	2.8	4.2	6.8	2.8	4.4
	Playing provocative game	10.8	2.4	7.1	9.6	6.0	8.5
	Cybersecurity metaphors in game	**14.1**	3.0	10.7	**14.8**	9.2	**11.9**
	Reality session	13.8	3.9	8.5	14.0	6.1	**11.6**
Day 2	Reality session idea exchange	12.1	4.1	8.6	12.5	6.5	9.2
	Meaning session	12.8	4.1	8.9	13.7	7.3	**11.0**
	Play session	14.0	3.8	8.5	14.5	6.8	**11.5**
Day 3	SGD Document & Presentation	12.3	3.6	10.3	12.9	6.2	9.2
	Serious Game Loop Exercise	12.5	3.5	9.4	14.5	4.9	8.3
	Serious Game Loop Design	12.5	4.5	**11.8**	12.5	5.5	9.3
	Prototyping session	**14.1**	4.9	10.5	13.6	7.3	9.2
Day 4	Playtesting session	12.4	4.3	7.5	**14.9**	6.4	8.5
	Prototyping development	12.7	4.9	9.4	12.7	6.8	10.0
	Development teamwork	12.7	4.9	9.4	14.0	6.8	10.6
Day 5	**Development - Part 1**	**16.3**	6.4	**11.9**	**15.8**	8.6	9.0
	Development - Part 2	**17.1**	6.0	**12.7**	**14.9**	**11.3**	9.3
Day 6	**Development – Part 3**	**15.3**	4.4	**12.0**	**14.7**	8.4	**11.3**
	Deliverables Preparation	**14.4**	4.6	11.6	14.7	6.8	10.8
	Peer assessment Games	13.5	7.0	9.7	15.2	8.3	9.9
	Closing Group Presentations	13.8	4.3	10.5	14.7	6.7	8.9

on Days 2,3,4, and 6 of the SSGJ. The IMI sub-scale scores of each individual participant support the findings in Table 3, with Figure 4 visualizing very high scores on the sub-scales of the IMI per day.

Table 4. Average IMI scores for each subscale per day of the SSGJ,with very positive scores of 5.50 and over highlighted in bold.

Day	Interest/ Enjoyment	Competence	Choice	Value/ Usefulness
Pre-SSGJ	**5.71**	4.29	**5.66**	**6.11**
Day 1	**5.55**	4.37	**5.67**	**6.14**
Day 2	**5.54**	4.91	**5.70**	**5.85**
Day 3	5.05	4.72	4.79	5.41
Day 4	5.29	5.00	4.88	**6.17**
Day 5	5.00	4.39	4.88	5.48
Day 6	**5.60**	5.09	**5.52**	5.40

It shows very positive results with regards to Interest/Enjoyment, Perceived Choice, and Perceived Value/Usefulness. The SSGJ was rated very highly on Interest/Enjoyment by 50%-61.5% of participants for 4 out of 6 days, as well as

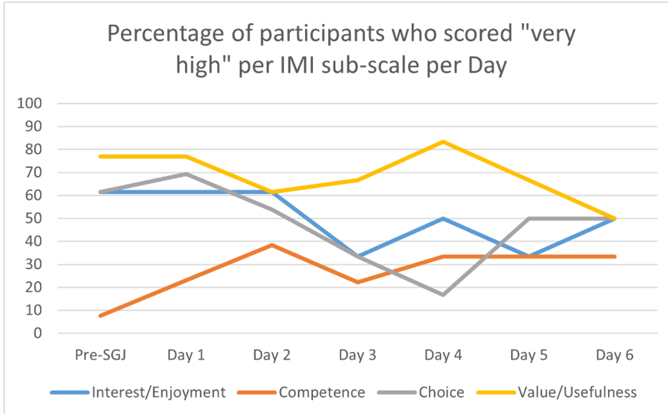

Fig. 4. Percentage of participants who scored "Very high" (6-7 out of 7) per IMI sub-scale per day.

perceived choice (50%-69.2%). Perceived Value/Usefulness is rated very high for all days of the SSGJ by 50%-83.3% of participants. The percentage of participants who rate Interest/Enjoyment very highly dips below 50% on Day 3 and 5, as well as Day 3 and 4 for Perceived Choice, but those sub-scales do stay positive on average. These days are the first day of serious game loop design and prototyping (Day 3), and the first days of creation and development (Day 4 & 5). The perceived Value/Usefulness of the activities of the SSGJ on those days remained very high though and even peaked on Day 4 to 83.3%.

4.5 Cybersecurity, Learning Mechanincs and Game Mechanics Cards

The questionnaires for each deck of cards consisted of twelve 7-point Likert-scale questions, three questions with tick boxes allowing participants to select multiple pre-defined options, and four free-text responses to open questions for further clarification. The response rates were 85% (i.e. 11 out of 13 participants) for the Cybersecurity Cards and GM Cards, and 77% (i.e. 10 out of 13 participants) for the LM Cards Questionnaire.

4.5.1 Cybersecurity Cards.

The cybersecurity cards provided a knowledge base for cybersecurity (see examples in Figure 1). Participants reported that they provided knowledge about individual cybersecurity concepts (90.9%), the wide scope of cybersecurity concepts (81.8%), the relationship between vulnerabilities, attacks and defences (90.9%), and terminology (63.6%). Participants reported they also provided a means for independent learning (72.7%) and self-efficacy by providing access to cybersecurity knowledge when the cybersecurity experts were not present (81.8%). They improved accessibility by acting as an interface to discuss cybersecurity topics with cybersecurity experts (81.8%) and others (63.6%) throughout the SSGJ. There were 63.6% of participants who indicated they provided inspiration for the design of the serious game. Furthermore, throughout the serious game design (63.6%) and serious game creation and development (72.7%), they were used as a reference point or reminder for ensuring the serious game stayed grounded in real cybersecurity issues (see Figure 5). The activities during which they were consulted the most were the discussion of the metaphors in the provocative game (54.5%), during the Reality session (45.5%) and Meaning session (54.5%) of TGD and during prototyping (54.5%).

4.5.2 Learning Mechanics Cards.

Participants agreed that the LM cards provided knowledge about individual LM concepts (90%), knowledge about the scope of LM concepts (90%), the relationship between LM concepts such as Analyse and Evaluate (80%), and

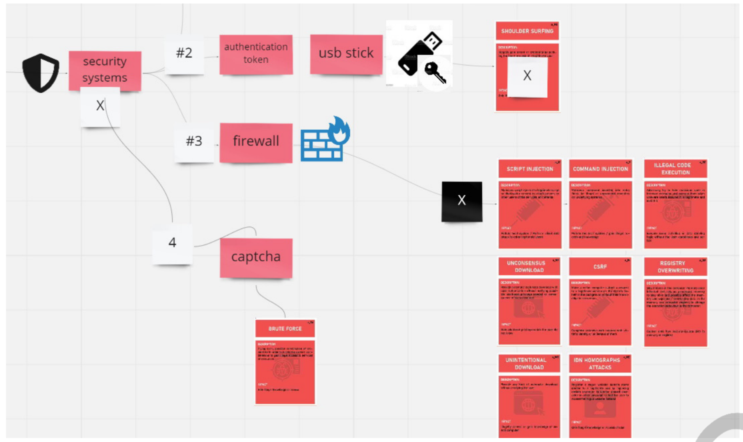

Fig. 5. Screenshot of part of a Serious Game Loop on Miro board, with Cybersecurity cards explicitly linked to elements in the game, such as a password-protected USB stick being vulnerable to a Shoulder Surfing attack.

terminology (80%). They also provided participants with a means for independent learning (90%) and self-efficacy by providing access to learning mechanics knowledge when the serious game design experts were not present during the SSGJ (80%). They improved accessibility, by acting as an interface enabling participants to discuss learning mechanic topics with serious game design experts (80%) and other experts or participants (77.7%). Similar to the cybersecurity cards, 80% of participants indicated they assisted in the design of their serious game. In particular during the Meaning session, design of the Serious Game Loop, and serious game creation and development (77.7%), they were used as a reference point or reminder for ensuring the linkage between their selected learning mechanics and the serious game they were creating (Figure 6). The activities during the Slow Gam Jam when the LM cards were consulted the most, were the Meaning session of TGD where the focus is on the educational part of the serious game, and the Creation of the Serious Game Loop based on the LM-GM map (both 60%).

4.5.3 Game Mechanics Cards.

Participants agreed that the GM cards provided knowledge about individual GM concepts (91%), the scope (91%), and the relationship between GM concepts (72.8%), and terminology (91%). More than half of the participants agreed that the GM cards were useful for independent learning (54.6%) and self-efficacy by providing access to game mechanics knowledge during the SSGJ when the serious game design experts were not present (54.6%), They improved accessibility, by acting as an interface for discussion with game design experts (72.8%) and other experts or participants (70%) throughout the SSGJ. They provided inspiration (63.6%) and acted as a reminder to link and ground the serious game in the selected game mechanics during game design (63.6%) and development (60%). The GM cards were consulted for different phase and activities throughout the SSGJ. In particular when "Discussing metaphors in the provocative game" (72.7%), all three stages of TGD (54.5%), "Creation of Serious Game Loops" (63.6%), Prototyping (45.5%), and Development (54.5%), covering all six days of the SSGJ (Figure 6).

4.5.4 Feedback on the Design of the Cybersecurity, Learning Mechanics, and Game Mechanics Cards.

Limitations of the design of each deck of cards are summarized in Table 5, indicating the percentage of participants who ticked that selection box with predefined options. For the design of each deck of cards, see Figures 1- 3. Table 5 shows that around half of the participants considered the total number of cards to be too high, the colour coding not clear, and the relationship between the different types of cards in the deck was not clear for the

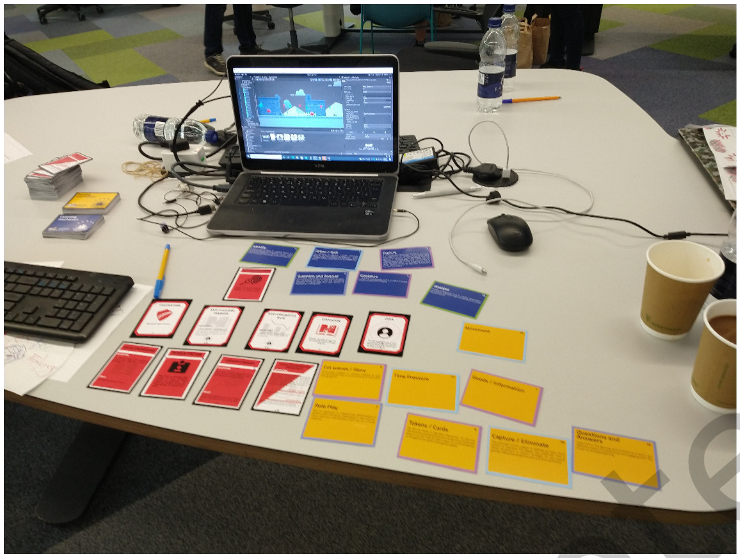

Fig. 6. Sub-selection of physical Cybersecurity cards, Game Mechanics cards, and LM cards being used as a reference point or reminder for ensuring the linkage with the serious game being created.

Cybersecurity cards and LM cards. The free-text option also indicated that almost half of the participants felt the Cybersecurity cards and LM cards limited their creativity for the design of their serious game.

Table 5. Overview of the percentages of participant selected limitations in the design of the Cybersecurity, LM, and GM cards.

Limitation	% for cybersecurity cards	% for LM cards	% for GM cards
Total number of cards is too high	45.50	50.00	27.30
Colour coding is not clear	45.50	30.00	18.20
Relationship between cards in the deck unclear	45.50	50.00	18.20

4.6 Assessment of Serious Games Using Cards

4.6.1 Self-Assessment.

Firstly, each team assessed their own game using the three decks of cards and discussed and explained their selection to at least one of the experts (see Figure 7). These discussions with the experts reflected an evolving understanding of both cybersecurity as well as serious game design, as participants could quickly, confidently, and adequately explain why certain cards were (or were not) part of their serious game design.

It also illustrated how many cards have been used the design of the serious game, which has been summarized in Table 6. It should be noted that a higher number of cards does not imply that the serious game design is better (or worse), but it does illustrate that the Cybersecurity cards played a prominent role in the serious game design.

4.6.2 Peer-Assessment.

The game was then played by another team, who subsequently also assessed it using the three decks of cards. A

Fig. 7. Participants selecting and discussing the Cybersecurity (red), LM (blue), and GM (yellow) cards used in their serious game design with an expert.

Table 6. Number of cards per serious game per team

Team	Cybersecurity	LM	GM	Total used
A	21	6	12	39
B	4	11	7	22
C	32	6	9	47
Total	57	23	28	108

member of the team that created the game would afterwards explain which cards matched in their opinion and which ones did not. The last column in Table 7 shows that, in comparison to the total number of cards used in the design, the number of cards selected that did not match the design according to the team who created it, is low. This indicates a high level of successful mapping of learning outcomes into the teams' games. It should be noted that it is not about a sub-selection of cards being "correct" or "incorrect", but about the discussion that occurred around this matching exercise, which provided valuable feedback from peers and experts on the design of the serious game, and how it was interpreted by others.

4.7 Engagement Between Scheduled Days of the SSGJ

The response rate for this questionnaire was 54% (i.e. 7 out of 13 participants). All participants indicated they had engaged with their serious game project in between scheduled days, as a: team (57%), subsection of a team (29%) and individually (29%). Responses indicated activities such as content creation for their game (86%) and for reflection on things they had learned (29%). Furthermore, further research was conducted in: cybersecurity (57%), learning context (43%) and on games (14%). Free-texts answers clarified the amount of work participants undertook outside scheduled days of the SSGJ differed depending on the phase: *"During this SSGJ, there was a*

Table 7. Number of cards in the peer-assessment of the serious game that matched the team's self-assessment

Team	Cybersecurity	LM	GM	Total	Did Not Match
A	9/21	5/6	5/12	19/39	4
B	0/4	3/11	2/7	5/22	3
C	12/32	5/6	8/9	25/47	2
Total	21/57	12/23	15/28	49/108	9

1-week time gap between Phase 1 day 1 and day 2. Being an introductory Phase 1 session there was not much to do during that week. So, while following the structured approach, selecting dates for the Serious Slow Game Jam are crucial. Also, there should be a one-day gap between Phase 3 day 5 and day 6. The last phase is a fast-paced event, and one extra day will help the participants a lot, to sum up and finalize their game development" – P4).

5 Serious Game Output From Serious Slow Game Jam

From the SSGJ, all three teams managed to deliver a playable prototype of their serious game as an output (i.e. one digital prototype and two non-digital prototypes), as well as any supporting documentation in the form of both a rule book for their game and a Serious Game Design Document (SGDD). In this paper, we will discuss one of the serious games that was further developed by serious game design experts, from the co-designed prototypes delivered through the SSGJ methodology. This game was decided to be further developed by consensus decision, chosen by experts in serious game design, cybersecurity, and software engineering (one external expert) in the SSGJ as the most fruitful for future development.

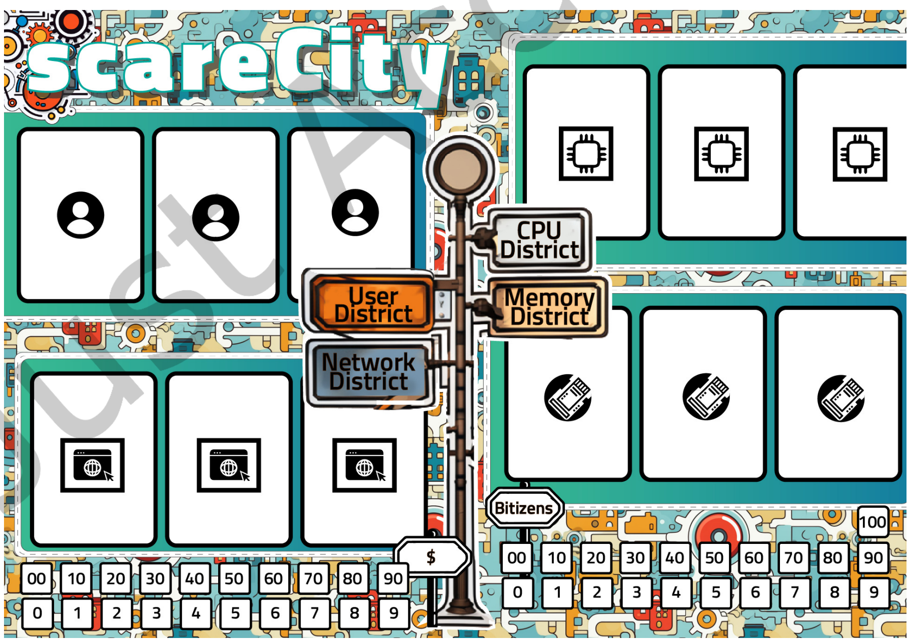

Fig. 8. The ScareCity game board showcasing the four program districts

5.1 ScareCity

ScareCity is a 2–4 player serious board game that is underpinned by the theme of secure software development lifecycles. This theme involves the integration of security into the software development lifecycle, coupled with additional processes such as risk analysis and security testing. In ScareCity, each player has their own game board (see Figure 8) and acts as the *mayor* of their city. The goal is to expand the city's features through *Program Cards* to manage both the security of the city and improve the trust of its *Bitizens*. Program cards can be abstracted to strategic security decision making (i.e., implementing defences), and fall into one of four *Districts*: CPU, Network, User or Memory. During the game, before the end of each round, *Impact Cards* are drawn which are akin to security attacks or defences, which either cause havoc to the player's program cards (causing them to degrade) or may bring benefits (i.e., increasing the number of bitizens). Examples of both *Program* and *Impact* cards can be seen in Figure 9. The first player to reach a total of 200 *Bitizens* wins the game.

Linking back at the theme of secure software development lifecycles, the goal of the game is to implement security into the player's city (programs), with aspects such as risk analysis and security testing done through analysing the available program cards to minimise the potential for serious impacts to a player's programs. The game also highlights the trade-offs that are often required to be made by cybersecurity experts due to limited resources. From this game design, it is clear that participant's in the Serious Slow Game Jam also successfully met the jam's intended learning outcomes regarding serious game design - understanding the importance of the three aspects of Triadic Game Design [41] evidenced by appropriate usage of reality, meaning and play elements during design.

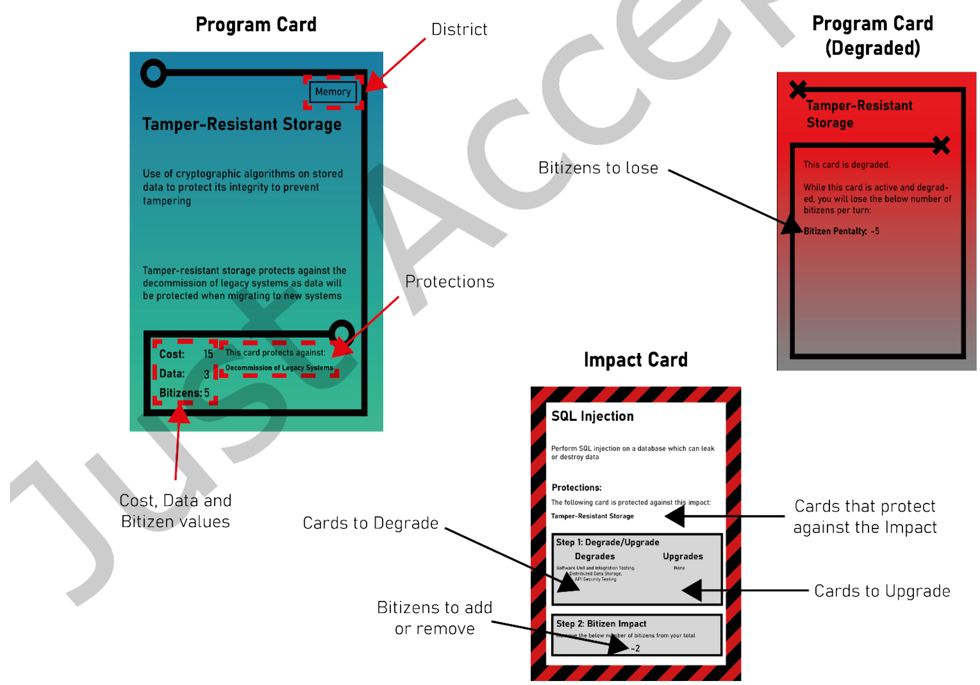

Fig. 9. Example of ScareCity Program and Impact Cards

The serious game ScareCity, including the game board, the Program- and Impact cards, and a detailed rule book, as well as other serious games produced as outputs of the SSGJ methodology, are all freely available from our Secrious project website (https://secrious.github.io/#seriousgames). It should be noted that a full, detailed evaluation of the serious games, that are produced as outputs of the SSGJs, with the target audience, is outside the scope of this paper, but is the focus of future work.

6 Discussion

This paper presents a first evaluation of the SSGJ methodology. The aim of this research is to evaluate how this SSGJ contributed to improving the understanding of cybersecurity for people who have little or no knowledge of cybersecurity. To investigate this, we were guided by the research questions in Section 1.1. We will now use the results to reflect on each research question in turn.

6.1 RQ1: How has the SSGJ affected participants' understanding of cybersecurity?

The SSGJ has contributed to improving participants' confidence in, and understanding of, cybersecurity, and insight into where their skills may be lacking. The assessment of the serious games using the three decks of cards presented in Section 4.6 showed that participants were able to assess their own serious game design using the cybersecurity cards. Discussion of their card selection with experts showed that they could effectively explain why certain cybersecurity cards were (or were not) part of their game design. Secondly, participants were able to communicate their knowledge of cybersecurity to others by relating cybersecurity concepts to elements in their game, as presented by the serious game output ScareCity in Section 5. Thirdly, participants were also able to successfully match cybersecurity cards to the serious game design of another team of participants, as results in Table 7 in Section 4.6 show. Matching all cybersecurity cards would be difficult though, as it also depends on how abstract the in-game metaphor is, the interpretation of the participant, and metaphors can be interpreted in multiple ways, as interpretive flexibility is an element of serious game design [2, 88].

The SSGJ also contributed to confidence in key skills regarding cybersecurity. Increased confidence levels in skills has been shown to be particularly key for motivation, as it can drive future learning and have a positive impact on self-efficacy [9, 13, 61]. Instances where confidence decreased (e.g. updating and reviewing existing code) indicated areas where confidence may have been inflated or perhaps unrealistic, and the process of the SSGJ has allowed participants to reassess and find areas in which they may need improvement. This is supported by the literature, which shows that participants with less experience in cybersecurity were more willing to acknowledge their mistakes and lack of expertise in their skills and decision-making processes while managing threats in a cybersecurity game, compared to experienced cybersecurity experts [30]. Almost half of participants indicated they had learned a lot about cybersecurity, but also almost half indicated that they had not learned much about the topic and wanted to learn more. Therefore, the SSGJ provided a starting point and triggered curiosity to learn more about cybersecurity. The divide between the participants was due to the skills and expectations of individual participants prior to the SSGJ [3], and indicated that the balance between game design and learning about cybersecurity, in particular for some of the activities in the last phase of this specific SSGJ, could be improved.

In addition, the results of the pre-/post-test in Section 4.2 indicated not only an improved confidence in participants' knowledge and understanding of the application domain of cybersecurity, but also an improved confidence in participants' knowledge and understanding of (serious) game design (from 12.5% to 75%). This was supported by the participants' ability to assess their own serious game and assess another team's serious game using the three decks of cards as presented in Section 4.6. This indicates that for the participants in the SSGJ, who had no or limited experience in game design and development, the SSGJ contributed to a better understanding of (serious) game design and development. This is in line with findings by [8], who found that learning the process of developing a game is one of the main benefits of a game jam. This is an important finding, as for a serious game design to be effective, there needs to be a balance between the three TGD constructs of Reality, Meaning, and Play [94].

A potential limitation of the one-group pre-/post-test results in the evaluation of this particular SSGJ, is that there is a possibility the difference in pre- and post test scores regarding participants' confidence in keys skills in cybersecurity and serious game design and development could partially have been caused by something other than participation in the SSGJ. This was mitigated by the qualitative evaluation of participants' knowledge and understanding of cybersecurity and serious game design and development during the SSGJ. This includes identifying cybersecurity metaphors in the provocative game and serious game design using TGD in Phase 1, the self assessment and discussion of the cybersecurity, LM and GM card selections with experts for their own serious game, and the peer assessment of another team's serious game in Phase 3, as can be seen in Table 1.

6.2 RQ2: How can the cards for the application domain (in our case Cybersecurity cards), Learning Mechanics cards, and Game Mechanics cards that are part of the SSGJ toolkit, assist in serious game design?

The intended role of the Cybersecurity cards, LM cards, and GM cards in the SSGJ toolkit, was to act as design tools and provide a knowledge base regarding the application domain and serious game design in the first phase of the SSGJ [3], and assist in the assessment of the serious games in the third phase of the SSGJ. However, regarding the design of the serious games, it was observed that the cybersecurity cards assisted in the design of the serious game in all three phases of the SSGJ, rather than participants' personal experiences with coding or with cybersecurity. As shown by the results in Section 4.5, the cards contributed to the SSGJ by providing a knowledge base for individual Cybersecurity, LM, and GM concepts and terminology, enabled independent learning and self-efficacy for when the experts were not present, and improved accessibility by acting as an interface for discussion. They also acted as a reminder to link and ground the serious game design in Cybersecurity, effectively mapped to LM and GM mechanics, which has been shown in the literature to be important in order to create an effective serious game [2, 40, 41, 54, 94]. However, the cybersecurity cards do not provide a complete overview of the domain and therefore could potentially limit the design of the serious games to the breadth provided by the deck. In fact, the cybersecurity cards and LM cards were sometimes challenging for participants to understand, and were both mentioned by participants as limiting their creativity for serious game design. This is unsurprising as a serious game needs to be grounded in the application domain and achieve the learning outcomes using learning mechanics, while an entertainment game does not have such restrictions. Furthermore, following the feedback from participants on the design of the Cybersecurity cards and LM- and GM cards presented in in Section 4.5.4, the design of those decks of cards will be improved accordingly [89].

6.3 RQ3: What are the Workload and Motivation levels of participants during the SSGJ?

In comparison to traditional, fast-paced game jams, which have a high workload and temporal demand [26, 35, 52], the SSGJ aimed to reduce the time pressure. Based on the NASA-TLX data presented in Table 3 it can be concluded that the SSGJ method has succeeded in this aspect. The mental demand is the highest for the three development activities on Days 5 and 6 (MD = 15.3-17.1), which is supported by free-text clarifications indicating participants found the developments activities the most challenging. This is not a surprise, as developing a game is an inherently high-mental workload task [11]. Temporal demand (TD) was only considered "high" (11-17) during the Serious Game Loop design in phase 2 (Day 3) and during the development activities in phase 3 (Day 5 & 6) of the SSGJ, with the final deadline of the SSGJ in sight. This may partially be related to participants' inexperience with those activities, which is a useful insight from a planning perspective. It indicates that participants may need additional breaks and support during these sessions to offset the fatigue from increased workload. However, temporal demand was not high for 15 out of 20 SSGJ activities (see Table 3), meaning the SSGJ managed to reduce the constraint of time pressure for most of the SSGJ.

Motivation levels reported in Section 4.4 indicate the SSGJ managed to engage people in software security concepts. Average levels of the subscale Interest/Enjoyment (5.00-5.60), and Perceived Value/Usefulness (5.40-6.17) were positive for all days and phases of the SSGJ. Interest/Enjoyment was even very high on average for Days 1,2,6, and Perceived Value/Usefulness was very high for Days 1,2,4. Perceived Choice was also very high on Days 1,2,6, and above average on Days 3-5. The latter overlaps with the days that temporal demand was the highest (see Workload results in Section 4.3). This appears to be partially related to obligations participants had outside the SSGJ, as participants during the SSGJ reported having coursework deadlines around those days that they prioritized. This is supported by the highest number of participants (4 out of 13) who chose to participate online instead of in person which occurred on Day 3, as well as the number of participants (6 out of 13) who skipped (part of) the day on Day 4 and Day 5 (see Table 2). The average Perceived Value/Usefulness of the activities of the SSGJ on those days remained positive though, and the percentage of participants who ranked the Value/Usefuleness very high even peaked during that time period on Day 4 to 83.3%. This is important, as it was during these days of the SSGJ that the serious games started to take shape.

As identified in the literature in Section 2.2, there is a gap in establishing a universal method with which to evaluate game jams against their intended outcomes [80]. One of the aims of the SSGJ methodology was to engage participants in the application domain, and reduce the workload in particular in terms of time pressure. However, workload and motivation levels are typically not quantitatively and systematically evaluated for game jams in a way that enables comparison, but are often based on observations of game jam facilitators during the game jam, or qualitative questionnaires and interviews with game jam participants regarding their game jam experiences afterwards [17, 62, 75]. As a result, there is no baseline for traditional, fast-paced game jams to compare the results regarding workload levels and motivation levels of the SSGJ methodology to. By systematically and quantitatively evaluating the workload after each activity of the SSGJ using the NASA-TLX [38, 39], and the motivation levels at the end of each day of the SSGJ using the IMI [60, 85], a contribution of this research is that it provides other researchers with an evaluation protocol to measure workload and motivation levels for their game jams, and provides a benchmark for both workload and motivation levels in other types of game jams to be compared against.

The SSGJ presented in this paper was organised as a hybrid event due to timing and COVID-19 restrictions at the time. This SSGJ took place in Edinburgh (United Kingdom) over a period of 5 weeks in May and June 2022. Local government guidelines on the Covid-19 protocol [34], which provided guidance on reducing risks from transmitting Covid-19 were followed, and made it possible to organise in-person events. Participants were therefore able to choose if they wanted to participate in the SSGJ in person or synchronously to prioritize inclusivity and accessibility for our diverse target group [3]. It was observed that the online participants engaged less with their team members and the SSGJ as a whole, and were more likely to miss part of the scheduled day. These findings are in line with findings in the literature on online events which were conducted during the COVID-19 pandemic and reported low turnout rates and low commitment effort of participants (e.g. [31]). Online participants in game jams in particular have been shown to interact less with other participants, would prioritize obligations outside the game jam, and had a tendency to work alone [22]. It also hampered the progress of teams on days when too many team members attended online instead of in person (e.g. by hampering the natural flow of feedback or taking too long to respond to questions). We therefore recommend participation in person rather than online. In addition, we recommend splitting participants over fewer teams with four or more participants at the start of the SSGJ, to be more resilient to absences.

6.4 RQ4: How has the “slow” format of the SSGJ affected participant engagement?

The "slow" format of the SSGJ impacted the learning experience in several ways. Besides reducing time pressure, results in Section 4.7 show the “slow” format also encouraged engagement in-between scheduled days of the SSGJ. Participants engaged by actively creating content for their serious game, reflected on things learned during the SSGJ, and conducted further research in cybersecurity, learning context, or games. In addition, they indicated they would like to continue working on their serious game after the SSGJ had finished. However, spreading the SSGJ across multiple weeks also makes recruitment for a SSGJ challenging, as potential participants struggle to commit full-time (from 10:00-16:00 GMT) on all scheduled days. Coursework deadlines, work obligations, differences in time zones (for online participants) and childcare responsibilities were mentioned as reasons for this. The flexibility and accessibility provided by the SSGJ format made it possible in some instances to work around this as it enabled participants to make up for lost time and catch up with the rest of their team between scheduled SSGJ events. However, we recommended scheduling the days of the SSGJ in such a way that it fits the requirements of most participants outside the SSGJ. Finally, results in Section 4.7 particular to this specific SSGJ, show the “slow” aspect needs to be better balanced with the pace and perceived workload of the activities in various phases. The Days 1 and 2 in Phase 1 with the introduction to cybersecurity and TGD can be scheduled closer together, while more space for content creation and refinement is needed between Days 5 and 6 for the development of the serious game prototype in Phase 3.

6.5 Limitations and Reflections

This paper has presented a first evaluation of the Serious Slow Game Jam (SSGJ) methodology. A potential limitation is the relatively small sample size of 13 participants spread over three teams participating in one SSGJ. In addition, all participants were MSc conversion students in computer science (but not in cybersecurity). Therefore, although we believe the presented findings for the evaluation of this specific SSGJ hold true for the SSGJ methodology in general, additional SSGJs with participants from different demographics are needed to validate these findings [93]. Furthermore, it should be noted that although the response rates for the various questionnaires throughout the SSGJ was high (i.e. 54%-100%), given the small sample size of 13 participants this could sometimes translate into a low real count in terms of number of participants. However, all results from the questionnaires are based upon the responses of 10-13 out of 13 participants, except the SSGJ Experience Questionnaire, which was the last questionnaire that was administered, and was administered after the SSGJ had finished. It had a response rate of 54%, meaning it was returned by 7 out of 13 participants.

As outlined in Section 3.1.3, for each team of participants, the outcomes of the SSGJ include a playable (digital or non-digital) serious game prototype, an accompanying rule book, and a SGDD. Although the resulting co-designed serious game prototypes are playable, further refinement and development by the serious game design experts might be necessary, in particular if a non-digital serious game prototype is to be translated into a digital serious game. The SSGJ methodology aims to enhance participants’ knowledge and understanding of the application domain, as well as produce playable serious game prototypes as an output. However, if the aim is to simply produce serious games for a given application domain, other types of game jams might be more suitable. Finally, as already discussed in more detail by Abbott et al. [3], running a SSGJ involves a significant commitment of time and resources from the organizers and experts as well as the participants. Also the challenges for recruiting participants, especially from mixed backgrounds and facing different accessibility barriers, should not be underestimated. Therefore, it is recommended, mainly for logistical reasons, that the SSGJ should be targeted to, and integrated with, formal or informal educational or continuing professional development programmes [3].

7 Conclusion

This paper has presented a first evaluation of the Serious Slow Game Jam (SSGJ) methodology, by evaluating a SSGJ as a mechanism for co-designing serious games in the domain of cybersecurity. The aim was to evaluate how this SSGJ contributed to improving the understanding of cybersecurity. To this end, we engaged 13 participants for 6 days over a 5-week period, into a multidisciplinary SSGJ involving domain-specific, pedagogical, and game design knowledge, and encouraged engagement in-between scheduled days of the SSGJ. The SSGJ provided support and mentorship from game design experts and cybersecurity experts throughout, while at the same time retaining the co-creation and supportive ethos of the traditional, fast-paced game jam. The confidence of participants improved (from 12.5% to 62.5%), and they reported that the SSGJ experience improved their understanding of cybersecurity, specifically in terms of vulnerabilities, attacks, and defenses. Also the confidence of participants improved (from 12.5% to 75%) for serious game design and development. The findings and resulting discussion in this paper provide useful insights into how the different aspects of the SSGJ, including the different phases and activities of the SSGJ and the different elements of the SSGJ toolkit, have contributed to enhancing understanding of cybersecurity and game design. As the SSGJ is intended to be flexible and applicable across multiple domains, these insights can be useful to researchers in the wider HCI community who are interested in using SSGJs to co-create serious games to improve understanding in other application domains. The findings regarding the workload and motivation levels of the SSGJ also provided a baseline for workload and motivation levels for other (types of) game jams to be compared against. The SSGJ format worked well in engaging participants in between scheduled days of the SSGJ, but the schedule for upcoming SSGJs will be slightly modified to better balance it with the pace of the SSGJ activities. Looking forward, we recommend in-person participation for the SSGJ, as in the hybrid format there was less active contribution from online participants who were also more likely to drop out. We will continue to use Serious Slow Game Jams as a mechanism to co-design serious games to improve the understanding of different themes within cybersecurity, focusing on code security, API security, and the security lifecycle. We believe the presented findings not only hold true for the specific SSGJ evaluated in this paper, but hold true for the SSGJ methodology. Therefore, those SSGJs will be evaluated following the same evaluation methods and evaluation procedures for the SSGJ methodology as outlined in Sections 2.2 and 3.2 to validate the findings presented in this paper, and investigate the usefulness and suitability of the SSGJ methodology for different target demographics. This will also allow the investigation of the potential effect of various participant backgrounds and skill sets on the learning outcomes. In addition, exhibitions for different audiences (i.e. cybersecurity-, HCI- and game design experts as well as the general public) will be organized to showcase the SSGJ toolkit and outputs. During these exhibitions, the serious games that have been co-created during our SSGJs will be played by their target audience, to evaluate their effectiveness to improve the understanding of cybersecurity for people who have little or no knowledge of cybersecurity.

8 Acknowledgements

The reported work is funded by the Engineering and Physical Sciences Research Council (EPSRC) with EPSRC Reference: EP/T017511/1.

References

[1] Daisy Abbott, Olga Chatzifoti, and Joanne Craven. 2021. Serious Game Rapid Online Co-design to Facilitate Change Within Education. In *Games and Learning Alliance: 10th International Conference, GALA 2021, La Spezia, Italy, December 1–2, 2021, Proceedings 10*. Springer, 233–238.

[2] Daisy Abbott, Olga Chatzifoti, and Sandy Louchart. 2022. Provocative Games to Encourage Critical Reflection. In *ECGBL 2022 16th European Conference on Game-Based Learning*. Academic Conferences and Publishing Limited.

[3] Daisy Abbott, Olga Chatzifoti, Sandy Louchart, Jamie Ferguson, and Shenando Stals. 2023. Serious 'Slow' Game Jam: A Game Jam Model for Serious Game Design. *Proceedings of the International Conference on Game Jams, Hackathons and Game Creation Events (ICGJ'23)*

(2023).
[4] Yasemin Acar, Christian Stransky, Dominik Wermke, Michelle L Mazurek, and Sascha Fahl. 2017. Security developer studies with github users: Exploring a convenience sample. In *Thirteenth Symposium on Usable Privacy and Security*. 81–95.
[5] Adobe. 2022. Adobe Forms. https://www.adobe.com/sign/how-to/create-online-forms.html
[6] Megumi Aibara, Satoru Kawakami, and F Masakazu. 2022. Lessons learned from serious game jams Organized by DiGRA Japan. (2022).
[7] Sylvester Arnab, Theodore Lim, Maira B Carvalho, Francesco Bellotti, Sara De Freitas, Sandy Louchart, Neil Suttie, Riccardo Berta, and Alessandro De Gloria. 2015. Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology* 46, 2 (2015), 391–411.
[8] Ali Arya, Jeff Chastine, Jon Preston, and Allan Fowler. 2013. An international study on learning and process choices in the global game jam. *International Journal of Game-Based Learning (IJGBL)* 3, 4 (2013), 27–46.
[9] Riikka Aurava, Mikko Meriläinen, Ville Kankainen, and Jaakko Stenros. 2021. Game jams in general formal education. *International Journal of Child-Computer Interaction* 28 (2021), 100274.
[10] Francesco Bellotti, Bill Kapralos, Kiju Lee, Pablo Moreno-Ger, and Riccardo Berta. 2013. Assessment in and of serious games: an overview. *Advances in human-computer interaction* 2013 (2013).
[11] Jonathan Blow. 2004. Game Development: Harder Than You Think: Ten or twenty years ago it was all fun and games. Now it's blood, sweat, and code. *Queue* 1, 10 (2004), 28–37.
[12] Liz Boyle, Fiona Hancock, Matt Seeney, and Laz Allen. 2009. The implementation of team based assessment in serious games. In *2009 Conference in Games and Virtual Worlds for Serious Applications*. IEEE, 28–35.
[13] Aibhín Bray, Philip Byrne, and Michelle O'Kelly. 2020. A short instrument for measuring students' confidence with 'key skills'(sicks): Development, validation and initial results. *Thinking Skills and Creativity* 37 (2020), 100700.
[14] Mateus Camara, Marcio Oyamada, Daniel Boll, Davi Giacomel, Felipi Matozinho, Gustavo Nomelini, Igor Negrizoli, Leonardo Vanzin, Lucas Amorim, Lucas Sá, et al. 2021. A&B Game Jam-Enhancing the Online Game Jam Experience With Milestones. In *Sixth Annual International Conference on Game Jams, Hackathons, and Game Creation Events*. 69–73.
[15] Bonnie Chinh, Himanshu Zade, Abbas Ganji, and Cecilia Aragon. 2019. Ways of qualitative coding: A case study of four strategies for resolving disagreements. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.
[16] Lacey Colligan, Henry WW Potts, Chelsea T Finn, and Robert A Sinkin. 2015. Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. *International journal of medical informatics* 84, 7 (2015), 469–476.
[17] Leandro de Almeida Melo, Tiago HS Leite, Fábio Freire, Marcelo G Perin, Fernando Figueira Filho, Cleidson RB de Souza, and Any Caroliny D Batista. 2019. What Motivates Different People to Participate in Game Jams?. In *Anais Estendidos do XV Simpósio Brasileiro de Sistemas Colaborativos*. SBC, 135–140.
[18] Menno Deen, Robert Cercos, Alan Chatman, Amani Naseem, Regina Bernhaupt, Allan Fowler, Ben Schouten, and Florian Mueller. 2014. Game jam: [4 research]. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. 25–28.
[19] Daniel J Dubois and Giordano Tamburrelli. 2013. Understanding gamification mechanisms for software development. In *Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering*. 659–662.
[20] Pat Dugard and John Todman. 1995. Analysis of Pre-test-Post-test Control Group Designs in Educational Research. *Educational Psychology* 15, 2 (1995), 181–198. https://doi.org/10.1080/0144341950150207 arXiv:https://doi.org/10.1080/0144341950150207
[21] Richard Eberhardt. 2016. No one way to jam: game jams for creativity, learning, entertainment, and research. In *Proceedings of the International Conference on Game Jams, Hackathons, and Game Creation Events*. 34–37.
[22] Travis Faas, I-ching Liu, Lynn Dombrowski, and Andrew D Miller. 2019. Jam today, jam tomorrow: Learning in online game jams. *Proceedings of the ACM on Human-Computer Interaction* 3, GROUP (2019), 1–27.
[23] Samuel B Fee and Amanda M Holland-Minkley. 2010. Teaching computer science through problems, not solutions. *Computer Science Education* 20, 2 (2010), 129–144.
[24] Felix Fischer, Konstantin Böttinger, Huang Xiao, Christian Stransky, Yasemin Acar, Michael Backes, and Sascha Fahl. 2017. Stack overflow considered harmful? the impact of copy&paste on android application security. In *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE, 121–136.
[25] Allan Fowler. 2016. Informal stem learning in game jams, hackathons and game creation events. In *Proceedings of the International Conference on Game Jams, Hackathons, and Game Creation Events*. 38–41.
[26] Allan Fowler, Foaad Khosmood, and Ali Arya. 2013. The evolution and significance of the Global Game Jam. In *Proc. of the Foundations of Digital Games Conference*, Vol. 2013.
[27] Allan Fowler, Foaad Khosmood, Ali Arya, and Gorm Lai. 2013. The global game jam for teaching and learning. In *Proccedings of the 4th Annual Conference on Computing and Information Technology Research and Education New Zealand*. sn, 28–34.
[28] Allan Fowler, Gorm Lai, Foaad Khosmood, and Richard Hill. 2015. Trends in organizing philosophies of game jams and game hackathons. In *GJ Workshop. FDG2015*.

[29] Allan Fowler, Xuelei Ni, and Jon Preston. 2018. The pedagogical potential of game jams. In *Proceedings of the 19th annual SIG conference on information technology education*. 112–116.

[30] Sylvain Frey, Awais Rashid, Pauline Anthonysamy, Maria Pinto-Albuquerque, and Syed Asad Naqvi. 2017. The good, the bad and the ugly: a study of security decisions in a cyber-physical systems game. *IEEE Transactions on Software Engineering* 45, 5 (2017), 521–536.

[31] Theodoros Georgiou, Lynne Baillie, Olga Chatzifoti, and Sheung Chi Chan. 2023. Future forums: A methodology for exploring, gamifying, and raising security awareness of code-citizens. *International Journal of Human-Computer Studies* 169 (2023), 102930.

[32] Hans Giessen. 2015. Serious games effects: an overview. *Procedia-Social and Behavioral Sciences* 174 (2015), 2240–2244.

[33] William Goddard, Richard Byrne, and Florian'Floyd' Mueller. 2014. Playful game jams: guidelines for designed outcomes. In *Proceedings of the 2014 conference on interactive entertainment*. 1–10.

[34] Scottish Government. 2022. Coronavirus (COVID-19): staying well and protecting others. https://www.gov.scot/publications/coronavirus-covid-19-staying-safe-and-protecting-others/

[35] Lindsay Grace. 2016. Deciphering hackathons and game jams through play. In *Proceedings of the international conference on game jams, hackathons, and game creation events*. 42–45.

[36] Frédéric Guay, Robert J Vallerand, and Céline Blanchard. 2000. On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and emotion* 24, 3 (2000), 175–213.

[37] David Gundry and Sebastian Deterding. 2022. Trading Accuracy for Enjoyment? Data Quality and Player Experience in Data Collection Games. In *CHI Conference on Human Factors in Computing Systems*. 1–14.

[38] Sandra Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[39] Sandra Hart and Lowell Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[40] Casper Harteveld. 2010. Triadic game evaluation: A framework for assessing games with a serious purpose. In *Workshop of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*.

[41] Casper Harteveld. 2011. *Triadic game design: Balancing reality, meaning and play*. Springer Science & Business Media.

[42] Casper Harteveld. 2012. *Making sense of virtual risks: A quasi-experimental investigation into game-based training*. Vol. 11. IOS Press.

[43] Chris Hecker and Jonathan Blow. 2012. The Depth Jam. https://www.chrishecker.com/The_Depth_Jam

[44] Hyunji Hong, Seunghoon Woo, and Heejo Lee. 2021. Dicos: Discovering Insecure Code Snippets from Stack Overflow Posts by Leveraging User Discussions. In *Annual Computer Security Applications Conference* (Virtual Event, USA) *(ACSAC '21)*. Association for Computing Machinery, New York, NY, USA, 194–206. https://doi.org/10.1145/3485832.3488026

[45] Discord Inc. 2023. *Imagine a Place.* http://discord.com

[46] Dominic Kao, Rabindra Ratan, Christos Mousas, Amogh Joshi, and Edward F Melcer. 2022. Audio Matters Too: How Audial Avatar Customization Enhances Visual Avatar Customization. In *CHI Conference on Human Factors in Computing Systems*. 1–27.

[47] Menelaos Katsantonis, Isavella Kotini, Panayotis Fouliras, and Ioannis Mavridis. 2019. Conceptual framework for developing cyber security serious games. In *2019 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 872–881.

[48] Annakaisa Kultima. 2015. Defining Game Jam.. In *Proceedings of the 10th International Conference on the Foundations of Digital Games (FDG 2015)*.

[49] Annakaisa Kultima. 2021. Game jam natives? The rise of the game jam era in game development cultures. In *Sixth Annual International Conference on Game Jams, Hackathons, and Game Creation Events*. 22–28.

[50] Annakaisa Kultima, Laura Piispanen, and Miikka Junnila. 2021. Quantum Game Jam–Making Games with Quantum Physicists. In *Proceedings of the 24th International Academic Mindtrek Conference*. 134–144.

[51] Gorm Lai and Foaad Khosmood. 2022. What Is a Game Jam? In *Game Jams–History, Technology, and Organisation*. Springer, 1–20.

[52] Gorm Lai, Annakaisa Kultima, Foaad Khosmood, Johanna Pirker, Allan Fowler, Ilaria Vecchi, William Latham, and Frederic Fol Leymarie. 2021. Two decades of game jams. In *Sixth Annual International Conference on Game Jams, Hackathons, and Game Creation Events*. 1–11.

[53] Nicole Lane and Nathan Prestopnik. 2017. Diegetic Connectivity: Blending Work and Play with Storytelling in Serious Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Amsterdam, The Netherlands) *(CHI PLAY '17)*. Association for Computing Machinery, New York, NY, USA, 229–240. https://doi.org/10.1145/3116595.3116630

[54] Theodore Lim, Sandy Louchart, Neil Suttie, James Ritchie, Ruth Aylett, Ioana Stanescu, Ion Roceanu, Ivan Martinez-Ortiz, and Pablo Moreno-Ger. 2013. Strategies for effective digital games development and implementation. In *Cases on digital game-based learning: Methods, models, and strategies*. IGI Global, 168–198.

[55] Shengmei Liu, Mark Claypool, Atsuo Kuwahara, Jamie Sherman, and James J Scovell. 2021. Lower is better? The effects of local latencies on competitive first-person shooter game players. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.

[56] Ryan Locke, Lynn Parker, Dayna Galloway, and Robin JS Sloan. 2015. The game jam movement: disruption, performance and artwork. In *2015 workshop on game jams, hackathons and game creation events (co-located with FDG-2015)*.

[57] Manuel Maarek, Sandy Louchart, Léon McGregor, and Ross McMenemy. 2019. Co-created design of a serious game investigation into developer-centred security. In *International Conference on Games and Learning Alliance*. Springer, 221–231.

[58] Manuel Maarek, Léon McGregor, Sandy Louchart, and Ross McMenemy. 2019. How could serious games support secure programming? Designing a study replication and intervention. In *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 139–148.

[59] R de A Maurício, Lucas Veado, Renata Teles Moreira, Eduardo Figueiredo, and Heitor Costa. 2018. A systematic mapping study on game-related methods for software engineering education. *Information and software technology* 95 (2018), 201–218.

[60] Edward McAuley, Terry Duncan, and Vance V Tammen. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport* 60, 1 (1989), 48–58.

[61] Léon McGregor and Manuel Maarek. 2020. Software Testing as Medium for Peer Feedback. In *United Kingdom & Ireland Computing Education Research conference*. 66–72.

[62] Mikko Meriläinen and Riikka Aurava. 2018. Internal barriers to entry for first-time participants in the Global Game Jam. In *Proceedings of the 12th European Conference on Games Based Learning*. 414–421.

[63] Mikko Meriläinen, Riikka Aurava, Annakaisa Kultima, and Jaakko Stenros. 2020. Game jams for learning and teaching: a review. *International Journal of Game-Based Learning (IJGBL)* 10, 2 (2020), 54–71.

[64] David R Michael and Sandra L Chen. 2005. *Serious games: Games that educate, train, and inform.* Muska & Lipman/Premier-Trade.

[65] Microsoft. 2022. *Microsoft Forms - Easily Create Surveys, Quizzes and Polls.* http://forms.office.com

[66] Miro. 2023. *Take Ideas from Better to Best.* https://miro.com

[67] Ine Mols, Elise van den Hoven, and Berry Eggen. 2016. Informing Design for Reflection: An Overview of Current Everyday Practices. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (Gothenburg, Sweden) *(NordiCHI '16)*. Association for Computing Machinery, New York, NY, USA, Article 21, 10 pages. https://doi.org/10.1145/2971485.2971494

[68] Juergen Musil, Angelika Schweda, Dietmar Winkler, and Stefan Biffl. 2010. Synthesized essence: what game jams teach about prototyping of new software products. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering-Volume 2*. 183–186.

[69] University of Bristol Bristol Cyber Security Group. 2020. CyBOK. https://www.cybok.org/

[70] Jeanette Falk Olesen and Kim Halskov. 2018. The dynamic design space during a game jam. In *Proceedings of the 22nd International Academic Mindtrek Conference*. 30–38.

[71] James Parker, Michael Hicks, Andrew Ruef, Michelle L Mazurek, Dave Levin, Daniel Votipka, Piotr Mardziel, and Kelsey R Fulton. 2020. Build it, break it, fix it: Contesting secure development. *ACM Transactions on Privacy and Security (TOPS)* 23, 2 (2020), 1–36.

[72] Oscar Pedreira, Félix García, Nieves Brisaboa, and Mario Piattini. 2015. Gamification in software engineering–A systematic mapping. *Information and software technology* 57 (2015), 157–168.

[73] David A Plecher, Christian Eichhorn, Janosch Kindl, Stefan Kreisig, Monika Wintergerst, and Gudrun Klinker. 2018. Dragon tale-a serious game for learning japanese kanji. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. 577–583.

[74] Atyanti Dyah Prabaswari, Chancard Basumerda, and Bagus Wahyu Utomo. 2019. The mental workload analysis of staff in study program of private educational organization. In *IOP Conference Series: Materials Science and Engineering*, Vol. 528. IOP Publishing, 012018.

[75] Jon Preston, Jeff Chastine, Casey O'Donnell, Tony Tseng, and Blair MacIntyre. 2012. Game jams: Community, motivations, and learning among jammers. *International Journal of Game-Based Learning (IJGBL)* 2, 3 (2012), 51–70.

[76] Jon A Preston. 2014. Serious game development: Case study of the 2013 CDC games for health game jam. In *Proceedings of the 2014 ACM international workshop on serious games*. 39–43.

[77] Argenis Ramirez Gomez and Hans Gellersen. 2020. More than looking: using eye movements behind the eyelids as a new game mechanic. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 362–373.

[78] Angela Ramnarine-Rieks. 2013. Learning by game design: Exploring its potential in undergraduate information literacy instruction. *Proceedings of the American Society for Information Science and Technology* 50, 1 (2013), 1–4.

[79] Romana Ramzan and Andrew Reid. 2016. The importance of game jams in serious games. In *10th European Conference on Games Based Learning: ECGBL 2016*. Academic Conferences and Publishing International Ltd.(ACPI), 538–546.

[80] Andrew J Reid, Phillip Smy, and Iain Donald. 2020. A Theoretical Framework for Game Jams in Applied Contexts. *2020 DiGRA international conference: play everywhere* (2020).

[81] Gary Reid and Thomas Nygren. 1988. The subjective workload assessment technique: A scaling procedure for measuring mental workload. In *Advances in psychology*. Vol. 52. Elsevier, 185–218.

[82] Hanna-Riikka Roinea, Mikko Merilainen, and Ville Kankainen. 2021. Jamming the assessment: The viability of a'Twine'game jam as a learning evaluation tool in higher education. *Journal of Play in Adulthood* 3, 2 (2021), 96–112.

[83] José Miguel Rojas, Thomas D White, Benjamin S Clegg, and Gordon Fraser. 2017. Code defenders: crowdsourcing effective tests and subtle mutants with a mutation testing game. In *2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE)*. IEEE, 677–688.

[84] Susana Rubio, Eva Díaz, Jesús Martín, and José M Puente. 2004. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied psychology* 53, 1 (2004), 61–86.
[85] Richard Ryan. 1982. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology* 43, 3 (1982), 450.
[86] Richard Ryan and Edward Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55, 1 (2000), 68–78.
[87] Philip M Sadler and Eddie Good. 2006. The impact of self-and peer-grading on student learning. *Educational assessment* 11, 1 (2006), 1–31.
[88] Phoebe Sengers, Kirsten Boehner, and Shay David. 2005. Joseph 'Jofish'Kaye.". *Reflective design. CC* 5 (2005), 49–58.
[89] Ryan Shah, Manuel Maarek, Shenando Stals, Lynne Baillie, Sheung Chi Chan, Robert Stewart, Hans-Wolfgang Loidl, and Olga Chatzifoti. 2024. Introducing and Interfacing Cybersecurity: A Cards Approach. *Computer Standards & Interfaces* (2024). https://doi.org/arXiv:2307.16535 Under Review.
[90] Valerie J Shute, Matthew Ventura, Malcolm Bauer, et al. 2009. Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In *Serious games*. Routledge, 317–343.
[91] Karina Kohl Silveira, Soraia Musse, Isabel H Manssour, Renata Vieira, and Rafael Prikladnicki. 2019. Confidence in programming skills: gender insights from StackOverflow developers survey. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*. IEEE, 234–235.
[92] Shenando Stals. 2024. *SECRIOUS Project*. https://doi.org/10.17605/OSF.IO/2MXDK
[93] Shenando Stals, Lynne Baillie, Ryan Shah, Jamie Iona Ferguson, and Manuel Maarek. 2025. Evaluating and validating the Serious Slow Game Jam methodology as a mechanism for co-designing serious games to improve understanding of cybersecurity for different demographics. *Computer Standards & Interfaces* 92 (2025), 103924. https://doi.org/10.1016/j.csi.2024.103924
[94] Giovanni Maria Troiano, Dylan Schouten, Michael Cassidy, Eli Tucker-Raymond, Gillian Puttick, and Casper Harteveld. 2020. All Good Things Come in Threes: Assessing Student-Designed Games via Triadic Game Design. *ACM International Conference Proceeding Series* 1 (2020). https://doi.org/10.1145/3402942.3403010
[95] Pamela Tsang and Velma Velazquez. 1996. Diagnosticity and multidimensional subjective workload ratings. *Ergonomics* 39, 3 (1996), 358–381.
[96] Stephen Uzor and Lynne Baillie. 2019. Recov-R: evaluation of a home-based tailored exergame system to reduce fall risk in seniors. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26, 4 (2019), 1–38.
[97] Jungpil Yoon, Seungwoo Lee, and Taiwoo Park. 2018. JediFlight: design and evaluation of wing-based flying experience in virtual reality. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. 309–320.
[98] Alexander Zook and Mark O Riedl. 2013. Game conceptualization and development processes in the global game jam. In *Workshop proceedings of the 8th international conference on the foundations of digital games*, Vol. 5.
[99] Michael Zyda. 2005. From visual simulation to virtual reality to games. *Computer* 38, 9 (2005), 25–32.

Received 21 June 2023; revised 30 September 2024; accepted 5 December 2024