Master's thesis

Software Engineering and ICT

2021

Jenny Lauronen

MATCHING STAKEHOLDERS TO THE SERIOUS GAME DEVELOPMENT CYCLE: A MIXED METHOD APPROACH

TURKU AMK
TURKU UNIVERSITY OF APPLIED SCIENCES

MASTER'S THESIS | ABSTRACT

TURKU UNIVERSITY OF APPLIED SCIENCES

Software Engineering and ICT

2021 | 84 pages, 2 appedices

Jenny Lauronen

MATCHING STAKEHOLDERS TO THE SERIOUS GAME DEVELOPMENT CYCLE

- A Mixed method approach

Developing a serious game is a challenging multidisciplinary task. The risk of creating a game that does not meet the learning goals is considerable. Balancing gameplay and instructional design to make a fun and effective learning game requires multiple experts across various development phases. Academic literature acknowledges the need to involve different target groups but does not specify when the experts' scarce resources would be best utilized. The availability of subject matter experts is particularly challenging, and the timing of their involvement requires careful consideration.

In this mixed-method research, a group of (n=32) key stakeholders – subject matter experts, target users, a pedagogue, and game developers tested a virtual reality (VR) fire extinguishing training application. The data for the study came from observing gameplay; collecting gameplay metrics; post-game questionnaires on user satisfaction, learning satisfaction and user experience; and focus group discussions.

The results convey the topics that different stakeholder groups pay attention to. The material revealed that subject-matter experts have to be involved in the early stages of the design process. Involving these experts in the later development stages is no longer beneficial because the changes they typically propose would be too costly. Target users do not have the necessary ability to recognize or communicate fidelity flaws and teaching approaches. They cannot be used in the creation of learning goals as they do not know what they should master. This research concludes with a serious game development cycle model stipulating the effective and efficient involvement timing of the key stakeholders.

KEYWORDS:

Virtual reality, serious game, virtual learning, serious game design cycle

OPINNÄYTETYÖ (YAMK) | TIIVISTELMÄ

TURUN AMMATTIKORKEAKOULU

Ohjelmistotekniikka ja ICT

2021 | 84 sivua, 2 liitettä

Jenny Lauronen

SIDOSRYHMIEN OSALLISTAMINEN HYÖTYPELIN KEHITYSKAARELLA

- Monimenetelmäinen lähestymistapa

Hyötypelin kehitys on haastava, moniammatillinen tehtävä; riskinä on kehittää sovellus, joka ei saavuta oppimistavoitteita. Pelattavuuden ja opetuksellisen sisällön yhdistäminen hauskaksi ja tehokkaaksi opetuspeliksi vaatii useiden eri asiantuntijoiden yhteistyötä kehitystyön eri vaiheissa. Akateeminen kirjallisuus tunnistaa eri sidosryhmien osallistamistarpeen muttei täsmennä kuinka niukkaa asiantuntijoiden työpanosta on parasta käyttää. Erityisen haastavaa on substanssiosaamisen mukaan saaminen, heidän osallistumisensa vaatiikin tarkkaa suunnittelua. Tässä monimenetelmällisessä tutkimuksessa, ryhmä (n=32) pelinkehityksen keskeisten sidosryhmien edustajia testasi virtuaalitodellisuus (VR) palonsammutus koulutussovellusta. Osallistujina olivat aihealueen asiantuntijat, pedagogi, kohderyhmän edustajat ja pelinkehittäjät. Data tutkimukseen kerättiin tarkkailemalla pelaajien käyttäytymistä pelitilanteessa; keräämällä pelimetriikkaa; mittaamalla kyselykaavakkeella oppimis- ja käyttäjätyytyväisyyttä sekä käyttäjäkokemusta; ja fokusryhmähaastatteluilla. Tulokset kertovat mihin eri sidosryhmien edustajat kiinnittävät huomionsa. Materiaali paljastaa, että aihealueen asiantuntijoiden kannattaa olla mukana kehitysprosessin alkuvaiheissa. Heidän osallistumisensa myöhemmin ei enää ole hyödyllistä koska muutokset, joita he tyypillisesti ehdottavat, tulisivat siinä vaiheessa liian kalliiksi. Kohderyhmän edustajilla ei ole taitoja tunnistaa ja kommunikoida todentuntuisuuteen tai opetustavoitteisiin liittyviä puutteita. Heitä ei myöskään voida hyödyntää opetustavoitteiden suunnittelussa, koska he eivät tiedä mitä heidän täytyy hallita. Tutkimuksen johtopäätöksenä on hyötypelien kehityskaari, joka sisältää tehokkaan tavan ajoittaa sidosryhmien osallistaminen.

ASIASANAT:

Virtuaalitodellisuus, hyötypeli, virtuaalinen oppiminen, hyötypelin kehityskaari

CONTENT

1 INTRODUCTION

Immersive virtual reality (VR) is becoming an important way to practice skills by doing. VR, utilizing head-mounted display and controllers for interaction simulates a virtual environment that immerses users to the extent that they have a feeling of being there (Bowman & McMahan 2007). VR can support or even replace expensive and limited real-world training facilities in many professional fields. It is becoming a considerable part of hands-on training, especially when practicing rare or dangerous situations that cannot be experienced or practiced enough in real life (Markopoulos & Lauronen 2019).

The potential of virtual reality is widely recognized (Bowman & McMahan 2007) and a myriad of VR-training applications are built for professional training. VR training applications can be seen as a serious game. "Serious games" refer to the use of entertainment game elements for purposes beyond mere entertainment, such as training, advertising, simulation, or education (Susi et al. 2007). The development of serious game is fundamentally different compared to games developed purely for entertainment. Despite the growing popularity of serious games and the benefits they offer, the development process has not matured into a clear process. The development of effective serious games is a time-consuming and challenging process, requiring an appropriate balance between game design and instructional design (Iuppa & Borst 2010). The challenge is to create beneficial and useful training scenarios with detailed environments for a realistic training experience. Overcoming the challenge of a proper development process requires a deep understanding of pedagogy, content matter and gaming (Braad et al. 2016).

There are currently no standard development tools intended explicitly for serious game design and development (Cowan & Kapralos 2017), (Braad et al. 2016). The risk of making a learning episode that does not meet specified learning goals is considerable. To meet the needs of effective serious game development, researchers have created models and frameworks. However, they often have a domain-specific or narrow approach. Most often, the frameworks

lack a learning-orientated approach (Ávila-Pesántez et al. 2017), (Dowidat et al. 2017), (Nacke et al. 2009), and (Braad et al. 2016). To focus on meaningful learning experience, it is helpful to understand how to bring in a particular subject matter. Collaboration between serious game stakeholders during the development cycle minimizes the risk of failing in achieving the learning goals (Kelly et al. 2007).

Various stakeholder groups are paramount in designing and developing the training scenario, testing usability and user experience, and validating the training during the development cycle. A broad range of experts, such as designers, developers, researchers, target users, subject matter experts, pedagogues and other stakeholders contribute to the process of realizing a learning experience. Subject matter experts and pedagogues are contextual experts. Designers, developers and researchers have a more technical approach. However, defining the goals and ensuring the game meets them, requires a process that organizes appropriate and well-timed stakeholder contributions to the exact development phases where they are required.

Although there is much literature on which stakeholder categories are required for successful serious game development, there is no concrete reporting on either the most efficient timing of their respective contributions or to what end each stakeholder group is actually useful (Braad et al. 2016). Literature vaguely describes that to build meaningful serious games, target user and content specialist involvement is needed and that meeting the learning goals is particularly challenging. The expectation is that specialist involvement is important when defining and evaluating learning outcomes. Target users, on the other hand, are best for testing learning and user experience. There is, however, consensus in that educational game development needs the subject matter expert (Olssen et al. 2011) to create a purposeful learning impact. The problem is that the availability of content experts is limited, and this work aims to find the most efficient way to involve them during the design and development of a training application in order to make a meaningful learning experience happen.

The research aims to presents a development life-cycle model that includes when and for what purpose different stakeholders should be utilized in creating serious games. The work leads to a purpose-oriented process of avoiding apparent design shortages by using professional insight from the beginning of the game development process. It arranges the stakeholder resources to their proper use, focusing on how the content specialists should be involved in the process. The process qualifies the optimal timing to use these experts and distinguishes when some other stakeholders could give similar input as the experts.

The scope of this thesis lies in the context of VR training and the objective is to determine where and when deep domain-specific skills and knowledge are needed when developing virtual reality training in order to fulfill its learning goals. A secondary objective lies in determining in which development phases experts can be replaced by another stakeholder category. The case study for this thesis is a fire extinguisher VR training application.

This study aims to determine the optimal involvement of different stakeholders in game development cycle by testing how Learning experience, User Experience, User Satisfaction testing results compare between four different stakeholder groups. The groups involved in the study are: (a) subject-matter specialists; (b) target users; (c) game developers; and (d) a game pedagogy expert, who was interviewed to ensure a complete picture of expertise.

The research followed a partially mixed concurrent dominant status methodology. The qualitative data primarily stemmed from focus group discussions and the pedagogy expert interview. This qualitative data was collected and analyzed following the constant comparison method (Boije 2002), and evaluated against the context of the Simulation Game Instructional System Design Model (Kirkley et al. 2005), pedagogy (Fowler 2015), and serious game design assessment framework (Mitgutsch & Alvarado 2012). The quantitative data came partly from a survey questionnaire, and partly from game metrics that were collected during gameplay. Due to a limited number of responses from certain stakeholder groups and a lack of questionnaire reliability, the quantitative statistics could not be

inferenced to a wider population. Therefore, the quantitative data is presented through various descriptive statistical techniques that include graphing and measures of centrality. Analysis revealed what the different tester groups pay attention to and what they speak about. This information is valuable when deciding on the right timed involvement of the stakeholder groups during the game development cycle.

The study is implemented by involving the stakeholder in gameplay testing, collecting data of their performance and having focus group discussions with them to compare their interests and focus. In the case study, the participants (n=31) presented: (a) experienced professionals as content specialists (n=21); (b) junior professionals as target users (n=3); and (c) game technologies engineering students who are in their final year studies and they presented a stakeholder group of game developers (n=7). All participants tested a beta-phase fire extinguishing VR-training application for professional training. The participants: (a) were observed during gameplay; (b) answered a questionnaire on user satisfaction, user experience and learning impact; and (c) participated in a focus group discussion to give feedback on their experiences. Additionally, the VR application collected metrics, offering information about activities during gameplay. The questionnaire and metrics present a quantitative part of the research and give the background to the focus group discussion with qualitative research objectives. In addition to this, the game pedagogue, a game development senior lecturer from Turku University of Applied sciences, was interviewed. The pedagogy was not part of the analysis itself, but the interview material was cross-referenced when analyzing the results.
This research generates a key understanding of exactly where we need domain-specific skills and what can be tested by other stakeholder groups. This research focuses on the serious game development cycle and does not attempt to validate or verify the application itself.

In the next section, the theoretical frameworks and models for serious game development are discussed, which leads into an unpacking of the problem that

this study attempts to solve. After this, the aims and objectives of the study are once more clarified. Section 2 introduces the theoretical background to the research question. Section 3 wraps up what is known and introduces the research question. Section 4 goes on to give a detailed explanation of the scientific approach used to solve the problem and answer the hypothesis. It explains the participants, experiment design and analysis methods in detail. After this, the research findings are presented before section 6 gives a detailed discussion of the analysis. Section 7 encapsulates the conclusions and recommendations. The references are listed in in section 8.

2 GAMES

Games are a particular part of culture and are one of the oldest forms of social interaction. Games are plays with formal structure, allowing people to go beyond imagination and direct physical activity. Game features include uncertain outcomes, agreed-upon rules, competition, different places and time, elements of fiction, elements of chance, prescribed goals, and personal enjoyment (Spanos 2021). The earliest board game pieces are in 5000 years old tombs in several places on earth. Modern chess rules took shape in Spain in the 16th century, and the first commercial board games date back to sometime in the 18th century. Outdoor games have also been very popular and are still played by all social classes around the world. Games are often viewed as the predecessors of modern sports. Throughout history, games have not only served social and hedonistic purposes, but also catered as instructional or learning tools. An example of an early learning game is the landlord's game, patented 1903 by Elisabeth Macie. The game was developed to demonstrate land grabbing, defined as very large-scale land acquisitions, with all its usual outcomes and consequences. The demonstration made it easier to understand and what can be done about it. The game was used in many variations to teach economy, and through time morphed into what we know today as Monopoly™. Miniature war games, used to simulate battle strategies, are another category of serious games that have been around since the 18th century—an example of this is a classic chess variant, called The Kings, dating back to 1780.

Games allow people to think strategically and simultaneously experience fortune, within the confines of the game's balance. People have a natural calling to play and as computing power emerged, it opened new dimensions to play with and video games became an interesting academic research tool. A series of games, generally simulating real-world board games, were created at various research institutions to explore programming, human–computer interaction, and computer algorithms (Smith 2014). Research results from digital game related studies paved the way to more entertainment-oriented gaming. The earliest electronic

game appeared in the US patent registration in 1947 and in the early 1970's, video arcade games were first offered to the public. The first commercially successful game was Pong—released on a game specific console. Video game consoles gained traction in the early 1980s, and the dominance of the industry shifted from the US to Japan. This same period saw the advent of personal computer (PC) games, specialized gaming home computers, early online gaming, and the introduction of LED handheld electronic games and eventually handheld video games (Rutter &Bryce 2006).

As technology progressed, several electronic games platforms came to market. Today, the market comprises primarily of PC, console, mobile, and virtual reality games—collectively referred to as videogames. The first generation of video games was text-based adventures, where players communicated by selecting various dialogue or action options to take one through the game. Game interaction then moved on to joysticks, controllers, keyboard-and-mouse, or motion-sensing device. This rapid development of device technologies has enabled an ever-increasing fidelity and acceleration in the intricacies of game rules and simulation, boosting the gaming industry to become the third biggest segment in the US entertainment market since 2018 (Gran view inc. Video Game Market Size, Share & Trends Analysis Report, 2020). Through the rising popularity of video games, there has been a natural uptake in using this media for training and other educational purposes.

Clark Abt was the first to formalize the idea of using games for purposes other than entertainment in his book 'Serious Games' in 1975 (Cowan & Kapralos 2017). "Serious games" usually refer to games used for 1) Knowledge or skill acquisition, 2) Motor functioning improvement, 3) Behavior change, 4) Generating awareness (in the case of advergames). (Ravyse et al. 2016). One of the foremost selling points of serious games for training purposes is that they allow the user to experience situations that are difficult or even impossible to achieve in reality (Lauronen et al. 2021). In the real world, dangers, rare events, costs, and ethical concerns limit the experiences required for effective and well-rounded training. Serious games-based training allows the training of a wide variety of skills—including analytical and spatial thinking, strategic planning,

recollection, psychomotor skills, and visual selective attention. Additionally, serious games have been shown to improve self-monitoring, problem recognition and solving, improved short- and long-term memory, and increase social skills (Mitchell & Savill-Smith 2004). Traditionally, serious games were dedicated primarily to PC and console platforms, but in the past three to five years they have found niche application on mobile and virtual reality head-mounted displays, too.

The term Virtual Reality (VR), was first used by Jaron Lanier, founder of VPL Research, in 1989, when he began to develop goggles and gloves, that were needed to experience what he called VR. VR devices and virtual environments have evolved over time to such an extent that people can undergo training in a virtual environment alone or in a shared environment with others. Virtual environments have many advantages over real ones, but most importantly, they can be used to experiment safely, they are controllable, and they offer repeat training without excessive additional costs. Furthermore, and budgetary constraints aside, any environment can be created, realistic or fictional, for the training.

VR affords experts an ideal space to replicate situations similar to those that they face, providing them the opportunity to train as often as necessary and experiment in a what-if fashion on the proper course of action but within an immersive, yet safe, three-dimensional environment. For example, a specific area in that VR has been used and researched thoroughly is training surgeons and medical students. VR simulations are being used for training, teaching, and planning for surgeries. It's an area that the advantages of virtual environments are most visible because it's easier and better to train on virtual bodies than real ones (Alaraj et al. 2011). Other environments where it is not feasible to train in real scenarios include firefighting or responding to bioterrorist attacks—such dangerous situations cannot be created in reality. VR environments, however, are capable of presenting these dangerous real-life emergency events so that target users can experience such chaotic and stressful crises and, through continuous training, be prepared to act accordingly if needed (Bailenson et al. 2008).

As with most emerging technologies, VR has had some challenges. The first lies in overcoming the novelty of the technology. Developers are keen to produce training environments to showcase the technologies without considering the pedagogical objectives. This presents a risk that many early VR training solutions might not fully demonstrate VR's potential learning impact. The technology itself has some usability limitations mainly centered around nausea and discomfort, and the controllers that are reported as unwieldy (Lauronen et al. 2020). In the early release phases of VR devices, the availability and costs of the systems have also limited their usage (Stavroulia 2018). The latest generation devices, however, has offered a remedy for many of the earlier shortages, thanks for better framerates and ergonomics. VR can offer now cheaper, scalable and measurable training. (Wohlgenannt et al. 2020). Controller issues are also being addressed through the development of hand-tracking in conjunction with haptic feedback gloves (Haptx, Senseglove, Manus, VRgluv).

Virtual reality (VR) head-mounted display (HMD) training environments is, therefore, becoming an affordable, measurable, and repeatable training alternative in a variety of disciplines (Sauders et al. 2019). VR has proven to be a good way to reduce training times, prevent errors, and even improve product quality. The motivation to implement this technology arises from the fact that it is presented as a low-cost alternative for training and preparation for the use of specialized equipment. VR offers the user an immersive, interactive, and innovative learning process (Naranjo et al. 2020).

To develop applications for VR training requires the use of a game engine and content creation tools such as 3D modelling and audio processing software, although the latter may be forgone by using ready-made assets from various content stores. Unity and Unreal engine are the most common engines to develop VR applications. Both are game engines that provide an intuitive and friendly programming environment resulting in VR application development that allows a high degree of immersion for its users (Naranjo et al. 2020). This thesis will use the VR acronym to represent a virtual environment using an HMD.

2.1 Key elements of VR learning environments

When creating learning applications in the 3D virtual environment, immersion and presence are always highlighted as a necessity to efficiently support learning. Immersion can bridge the technological, psychological, and pedagogical aspects of the environment. The concept of immersion builds on two other properties: representational fidelity and interaction (Fowler 2015). The concept of presence is a psychological state that arises from immersive systems – a sense of being there (alone or together) and a sense of presence. Immersion is important in skills learning to maintain the situational interest (Ravyse et al. 2017).

A very simplified but practical framework dealing with different aspects of immersion and skills learning, is offered by Mayers and Fowler (1999). It splits skill into three phases. The first step is the presentation of the concept to be learned. It can be a lecture, book, virtual familiarization, or all of these. The second step is that learners start to work with the concept. Work can be laboratory tests, questions, writing, manipulations – actions where the learner can have feedback, and the learner's actions control the flow of information. Here immersion is in the task. In this stage, skills training could take place with VR. The third step is applying the new concept to a broader social context – a dialogue. Here the learner must test their new understanding by exposing it to interaction. Immersion plays a significant role in each of the three phases. This thesis is particularly interested in the development of VR for the second phase and how it could transition learners to the application step.

As the training application's fidelity is an essential element for an immersive learning impact, a better understanding of the various fidelity elements becomes important. Ravyse et al. (2020) lists qualities of fidelity as presented in figure 1 and states that high fidelity (closely resembling real-life) promotes successful learning as long as the high fidelity is focused on the elements that confer learning. High realism outside of the topics to be learned will become a distraction and lower the learning results.

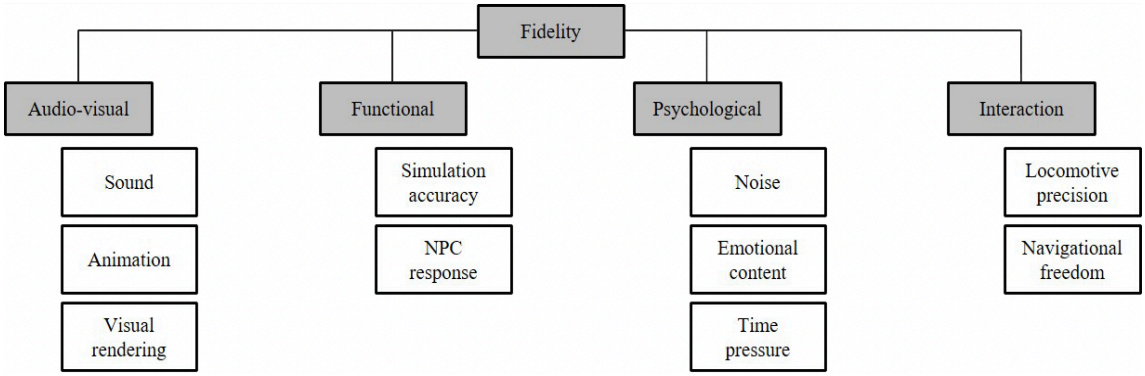


Figure 1: Fidelity organogram (Ravyse 2020).

2.2 Skills acquisition with serious games

Serious games are based on learning through virtual practice. The immersive nature of serious games improves performance and accelerates the acquisition of knowledge and competencies in the professional training process (Alverez & Djaouti 2010). Motivation and commitment caused by immersion and serious games fidelity improve professional performance among learners (Billet et al. 2014). The brain does not distinguish between natural and artificial stimuli (Allal-Chérif et al. 2016)

The competencies acquired by VR play can be transmitted and replicated in daily practices. Players can connect virtual and actual practices when the game is not too far from the work environment (Allal-Cherif et al. 2016). That is, when the context of the game is relatable to real working practice. Furthermore, virtual learning offers support for many learning styles and makes learning accessible in the areas that require visual, auditory, and kinesthetic engagement (Freina & Ott 2015). The advantage of VR training applications for this aspect is that training in a realistic environment can happen without the need for building physical setups. Despite the growing popularity of serious games and the benefits they afford, the development of effective serious games is a complex and time-consuming process, requiring an appropriate balance between game design and instructional design. It has been suggested that a lack of proper instructional design will lead to ineffective serious games (Cowan & Karpalos 2017).

2.2.1 The pedagogical frameworks of virtual learning

Learning specifications and "design for learning" as a perspective would increase the attainment of learning outcomes (Cowan & Karpalos 2017). Still, it's common that the pedagogical aspects of learning environments in virtual reality fall to secondary class—the excitement of novel technological possibilities takes precedence over sound pedagogy.

Learning design is a holistic activity of designing and planning activities as a part of a particular learning session, unifying technological and pedagogical aspects to support learning outcomes (Fowler 2015). Pedagogy exists to help inform the design and use of the educational VR applications. Intended learning outcomes guide learners in what they are expected to know, understand, and be able to do after completing a learning course. Often, learning outcomes are assumed to be defined, causing them to remain unclarified and not targeted appropriately in VR training applications (Biggs 2003).

Dalgarno and Lee (2010) identify five key task affordances that work toward the achievement of learning outcomes: Spatial knowledge representation, experimental learning, engagement, contextual learning, and collaborative learning. Some of them are generic and should always appear, and some, like spatial knowledge, only apply in some activities. More critical to virtual learning, Anderson et al. (2001) have recognized that a learning outcome (what is learned) is not the same as a learning activity (how it is learned).

The game as a learning activity has some specific aspects to consider (Malone & Lepper 1987) to keep learning fun and efficient. Digital game-based learning should have:

- learning goals that students find meaningful,
- multiple goal structures and scoring to give students feedback on their progress,
- numerous difficulty levels to adjust the game difficulty to learner skill,
- random elements of surprise,
- an emotionally appealing fantasy and metaphor that is related to game skills.

There are various methods and processes to integrate the serious content within the mechanics to balance fun and education and the high-level choices that identify and specify learning outcomes may already assume the virtual environment is a suitable media for teaching. In other words, how much learning takes place in a virtual environment may often be an assumption. The low-level choices, for instance, the cases used to achieve the learning goals, must be aligned with the target user (Ryanand & Charsky 2013) and the context in which they operate daily.

Fowler (2015) suggests establishing learning goals first and then choosing the learning experience according to the learning context. This opens possibilities to innovative new pedagogical environments and reduces the risk of replicating the real world's learning environment and modality. Fowler continues by claiming that the chance of inventing new, more efficient learning methods increases when focusing on actual learning goals. One should, however, not lose sight of the remaining serious game production phases and how the various stakeholders play a role in them.

2.2.2 The game development stakeholders

For a game designer, it's hard to foresee the contexts of the usage and although purposeful training is their intention, it is not necessarily in the designers' focus (Mitgutsch & Alvarado 2012). When we add new interactive technology, such as VR, where the interaction methods and the environment are new for the target group, the challenge expands. Design models that designers use as their primary toolset help, preferably together with the user (Ávila-Pesántez et al. 2017), to identify the mechanics, dynamics, and aesthetics of a training application. Designers are also often restricted because serious game target audiences tend to be very specific and smaller than entertainment game player groups, which leads to smaller budgets and strictly B2B (business to business) models.

One of the most trusted methods for involving target users is by means of a user-centric design methodology. This method involves target users throughout the design process and is helpful in pointing out user requirements (Fullerton 2014).

Yet, including only users in the design process is insufficient, particularly before there is a playable version of the game. Although end users provide learning outcome insight through gameplay, they cannot contribute to the actual domain knowledge during earlier serious game design phases. Such knowledge requires research and domain-specific expertise (Marchiori et al. 2012). When creating learning games, researchers, designers, and content experts work side by side in the design process, often even in many roles, and the design ideas and preferences differ, depending on the role. For example, the workload of creating application features is often underestimated, which brings resource challenges in the later phases of the development (Kelly et al. 2007). Serious game design models and frameworks are needed to define the game design process and should align the various phases with stakeholder inputs.

2.2.3 The serious game development cycle

There is not only one development cycle to follow, but the best and suitable model is case-specific (Braad et al. 2020). The development of a serious interactive game requires stages where the idea determines the concept and the concept continues to raw prototypes and more detailed design (alpha) and finally to wider beta testing, and eventually implementation and post-production with the commercial aspects (Figure 2). The model presented in figure 2 is a derived set of steps from various sources and can be viewed as an adaptation of the ADDIE model (analysis, design, development, implementation, and evaluation). This thesis merely uses the linear model to illustrate the distinct phases because, as with the original ADDIE model, it would be too rigid when using it in linear fashion for real-life situations (Braad 2020, Kirkley 2005, Naranjo 2020, Appelman 2005).

Figure 2: The phases of the serious game development life cycle.

To combat the rigid structure of a strictly chronological model, research and practice are clear that serious game development should happen in iterative cycles (Ramadan & Widyani 2013, McKenney & van den Akker 2005). Iterative cycles (as presented in figure 3) allow testing results and new information to surface in the final product. Moreover, iterations and testing rounds allow different development stakeholders, such as users and specialists in different areas, to participate in the process. Instructional designers and game designers must agree to use the same process to successfully create a balanced learning game. This thesis research is based on the Instructional System Design Model (ISDM) (Figure 3) (Kirkley 2005). The model foundation is on both instructional and game design models, ensuring the resulting game is instructionally sound. Instructional games are likely only one part of a larger learning environment. Therefore, the process needs to support the overall learning environment's design (Appelman 2005). Games are complex and interlinked environments where a simple change could significantly impact the narrative, negate meeting a learning objective, or interfere with planned performance assessment events (Kirkley 2005). As such, the ISDM also helps to safeguard the design process against the dynamic, often volatile, nature of game development.

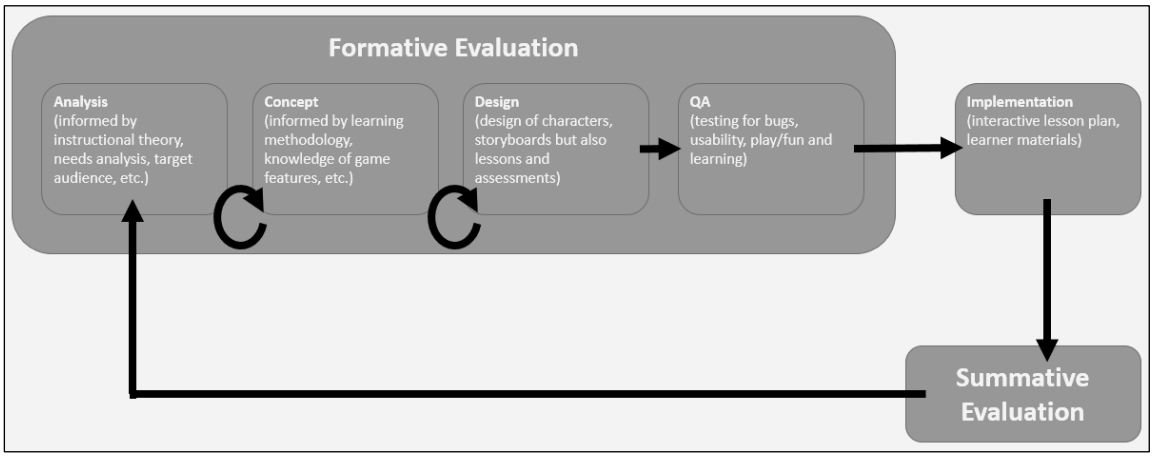


Figure 3: Simulation Game Instructional System Design Model (By Kirkley et al. 2005).

2.2.4 Testing serious games

No matter how precisely a serious game instructional design model is followed, there is a continuous requirement to test the game throughout the various phases of development. In broad terms, testing a serious game happens on two fronts: (a) testing the game and gameplay; and (b) assessing whether the game meets the learning outcomes.

For gameplay testing, Olsen et al. (2011) claim that issues in the area of usability should be rigorously tested. The procedure by Olsen et al. distinguishes the requirements unique to serious games, namely usability, playability, and learnability. The usability in educational games is tightly weaved with purposefulness and user experience. If usability (for example, controller difficulties) interferes with gameplay tasks, much attention will go into handling the controls and reading the instructions. This has the impact that learning results will suffer, or worse, that the learner will abandon the game.

Testing usability can be done in a variety of ways, from observation with think-out-loud protocols to questionnaire-type assessments. There are many different assessments in the form of questionnaires for usability testing. Some are standardized measure for general purpose testing, while others can be more company- or application-specific. Most often, they consist of Likert scales that focus on different aspects of usability. The usual usability elements to be tested are: (a) display characteristics, including the location of information on screen and legibility; (b) language usage; (c) the ease of interaction with the program and difficulty carrying out desired tasks; (d) how easily the system is learned; (e) general consistency and other subjective measures associated with how well the system operates.

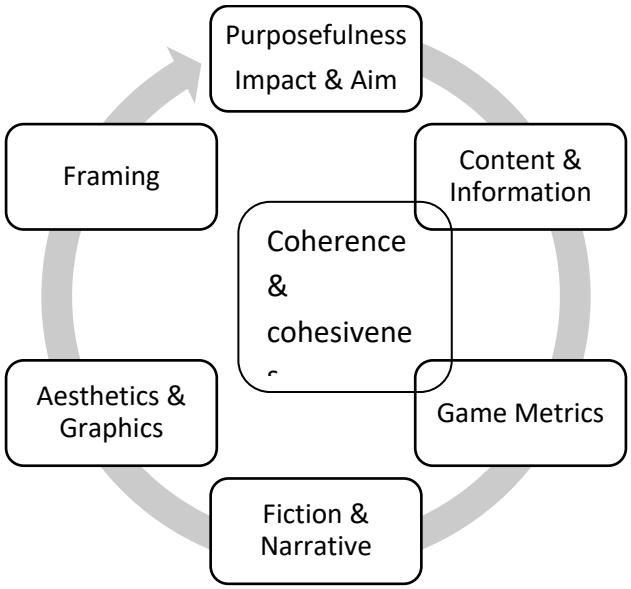
If usability is seen as technical efficiency, playability presents a broader scope of interaction functionalities and different tool integrations. Playability testing seeks the balance of fun and educational content. These include scales of immersion (or presence), flow, and engagement. Ibrahim (2020) presents the elements of playability as: (a) intrinsic; (b) mechanical; (c) interactive; (d) educational; (e) artistic; and (f) inter- and intrapersonal. Ibrahim suggests a test environment for

playability that includes a set of educators and different user profiles (experienced gamer -users to inexperienced). The test should have three steps: (a) heuristic evaluation of the defined playability problems; (b) a cognitive walkthrough, by playtesting and researcher observation; and (c) testing with real users, ensuring a high-quality playable game.

Learnability testing should be run in the same intervals as usability and playability to keep the learnability consistently high when proceeding in development. Learnability differs from learning outcomes assessment in that learnability measures how easily the users can familiarize with how the game work. Learning outcomes assessment refers to measuring how much knowledge or skill users have accumulated through gameplay —typically by means of a pre-test-post-test methodology.

2.2.5 The Serious games assessment framework

Assessing learning effectiveness of games is reflected through the Serious Games Assessment Framework (SGAF) (Mitgutsch & Alvarado 2012). The SGAF does not specify how to test at each level, or even who should be consulted in the testing, but rather identifies various key facets of serious game design where designers must not lose focus of the game's learning outcomes. Losing sight of the learning outcomes in any of these areas would jeopardize the purposefulness of the game. The SGAF suggests that evaluation criteria should be consistently used to cross-reference the game's purpose with the game design elements. The SGAF elements include purposefulness, content and information, game metrics, fiction and narrative, aesthetics & graphics and framing of the game system (Figure 4). The game's purpose should be built into and assessed within each of the elements.



Picture 4: Serious Games Assessment Framework (Mitgutsch & Alvarado 2012).

The coherence and cohesiveness of the game system encapsulates how the elements relate to each other and to the game's purpose. If cohesiveness is deficient, the system becomes conflicted and may reduce to less than the sum of its parts. A holistic design-related evaluation will bring to life a way to view the serious game's strengths and weaknesses. It also offers methods to keep an eye on the element coherence while designing the serious game. However, it does not instruct when and by whom the evaluation should occur.

In this thesis, the application usability elements (usability, playability and learnability), discussed in § 2.2.4, and learning effectiveness, discussed in § 2.2.5, are determined by means of a questionnaire that questions each participant's sense of user experience, satisfaction with the app and perceived learning impact. Through the questionnaire, it was hoped to establish if the participant groups differed in the way they considered each of the two serious game aspects, namely usability and learning effectiveness.

2.3 Serious game development phases and stakeholder summary

This section unpacks serious game development into a series of distinct phases to expound on the focus of each phase and describe the activities that are best undertaken in each of the phases to ensure a coherent and cohesive serious game application. The ideation, conceptualization, alpha 1, alpha 2, beta and final production phases are discussed from the basis of work done by Olsen (2011).

Ideation

Olssen et al. (2011) emphasize that it is essential to identify the target audience correctly. Subject matter experts can be helpful according to Olssen, but the actual users must be interviewed in pre-development phases so that critical details are not overlooked. Basic demographics of the target group are required as well as background info, such as prior knowledge, gaming experience and reading level. Other user capabilities and limitations, such as disabilities, may influence interaction with the game. The knowledge baseline and cognitive capacities are primary to decide over things like navigation and controls. Also, relevance and need, and knowing about possible resistance, helps make decisions on the approach, art, and communication. Cognitive loads, caused by for example, language or complex game structures, can significantly consume the achievement of learning outcomes and motivation and take the focus away from the topic.

Conceptualization

This phase focuses on creating storyboards of the concept. The storyboards stem from objectives, game features, implementation, and outcomes and are central to taking target user considerations into account. Storyboards should include game design, style, and art. When testing them with target users, it is important that storyboards progress just like the game would. This is usually a desktop exercise and becomes useful for checking the flow or progression and can be used for testing further ideas.

Prototyping

This phase should create a prototype of the game for a small group of the target users to play and give feedback on its usefulness, structure, and characteristics.

During this process, the questions should target in-game features like narrative, general usability, and ease of understanding. Usability issues here are general - whether the control scheme seems to fit the game and makes sense, whether the screen order and progression seem logical, and whether the objectives seem clear. The test can be done using questionnaires like the System Usability Scale (SUS), Questionnaire for User Interface Satisfaction (QUIS), or Technology Acceptance Model. The items relate to perceived usefulness, behavioral intention, ease of use, application-specific self-efficacy, enjoyment, opinion of game elements, general usability and playability, and player preferences. If a test is done using the questionnaires, follow-up discussions are essential to form a deeper understanding behind the measurement scales.

Building and testing Alpha 1

The very first version of the build should concentrate only on functionality. The testing covers the usability principles, functionality, and limitations of the game. This testing round, sometimes referred to as "game-breaking," can usually be conducted by in-house developers. Testing can be informal with the main aim of potential problem detection. Testers should record bugs and human factors like hard-to-read texts. After this, a small tester group of regular, healthy users from an easily accessible population can test, and the target is once again to catch prominent usability issues. Any of several methods work in this phase—observing and asking questions, think-aloud protocol, videoing, and so on.

In this stage, the learning outcome assessment and playability testing can be challenging due to bugs and usability issues are disturbing the performance. Readiness levels must be noticed when reviewing the results. SUS and QUIS can be used here, but it's recommendable to choose only the relevant areas of each questionnaire as too long surveys may distort the results. These surveys are, unfortunately, not designed with serious games in mind, and thus lack the greater depth desired when conducting gaming research. In many instances, relevant gaming usability questions are added. The questionnaire data should be analyzed into a usability report and bug tracker software is employed to follow the status of reported issues.

Build and testing Alpha 2

The new build, developed from Alpha 1, should have complete functionality and be entirely usable, and free from major bugs. Five individuals from the target group is a sufficient number of testers to ensure there are no population-specific usability issues, learning objectives are in place, and the game is useful. Testing proceeds in the same format as the previous round—only the playability and learning assessment scores are added if not assessed before. Also, more questions can be added to receive more feedback from the content and learnability side. The report is once more presented to the developers.

Build and testing of Beta

In this stage, the development team should have the entire spectrum of game experience in place. This includes the core of the game, the main implementation of art and all other game elements. Beta version testing requires a comprehensive assessment of the entire gaming interface and must, therefore, be tested with the target population. Since this phase is about detailed heuristic interface evaluation, a small group of five is sufficient for testing. The purpose of this testing is to ensure there are no lurking bugs or deeply hidden usability issues. The process can be identical to the former stages, but with a much more systematic approach.

Final Build

The final build is the completing phase, where the final draft of the game is prepared. Testing can be done in-house to ensure that the final draft is free of any bugs, that all issues have been fixed, and that all of the game's goals are met.

The game's evaluation is often done by self-reporting, observation, and interviews. For example, the gameplay experience questionnaire (GEQ) can be used. It's also typical to evaluate the serious game, comparing it with traditional teaching method effectiveness (Ávila-Pesántez et al. 2017). This is usually done by measuring target-user knowledge or skills gained through the use of the serious game in question and comparing it to another target user-group who underwent the traditional training.

Implementation

Implementation Adoption (Dowidat et al. 2017) consists of deploying the serious game in a comprehensive training protocol that is appropriate to each target user-group. Adoption can be facilitated by an internal marketing campaign and by involving a community manager to unite targeted populations around the game. Sophistication (Nacke et al. 2009) of the serious game is the evolution and continuous improvement of the game based on target user feedback, new standards, new tools, new professional competence, and environmental changes (particularly, those of a socio-economic nature). If needed, community functions can be established (Allal-Cherif et al. 2016).

Evaluation

The evaluation model suggested in this thesis is presented in the serious game assessment framework (§ 2.2.5). When the training applications have been used in real life circumstances with real users, refinement of the app should take place. Refinement comes as a result of combining the SGAF evaluation and the previous implementation phase sophistication outputs.

3 PROBLEM STATEMENT

It is a recurring theme that the target users should take the central role in serious game or application development. Theory suggests that target users must be used at every phase of development with a broad range of experts, such as designers, developers, researchers, target users, and other stakeholders, needing to be organized in the design process and they should work side by side in the design process. Also, the instructional designers and game designers must choose and use the same process to successfully create a balanced learning game. Although the related literature presents some course ideas on when to draw on various stakeholder groups, no concrete evidence exists about the most efficient manner to use the various skillsets required to develop an effective virtual training application. Table 1 presents what the studied material suggests.

Table 1: Summary of related work's suggestions how to involve stakeholders into serious game development cycle.

Serious game development cycle	**Role**	**Stakeholders**
Analysis and concept	Orientating and defining the system to build	Target users Domain experts
Prototyping	Validating the concept	Target users; different stakeholders working together in the design process
Testing alpha	Ensuring the playability and achievement of learning goals during the development	Target users; In-house developers
Testing beta	Ensuring the application is production ready	Target users
Assessment and evaluation	Assessing how the application meets the requirements	No mentions

This is the situation today. The learning games are popular but combining learning impact and playability is the challenge. The collaboration between the different stakeholders is suggested to be a solution, but no detailed answer on how to implement them throughout the game development cycle is available. This research will find how to involve different key-stake holders in the serious game development process. Also, since content specialists are the rarest resource, this thesis will lift out how to utilize this resource most effectively.

4 METHOD

This study made use of a VR application for fire extinguisher training and implemented quantitative and qualitative methods to collect data toward answering our research question. Mixed methods research represents research that involves collecting, analyzing, and interpreting quantitative and qualitative data to investigate the phenomenon. The mixed-method research was planned and implemented in a partially mixed concurrent dominant status (Leech & Onwuegbuzie 2009, Chun et al. 2019). This means that the research collected both qualitative and quantitative data at approximately the same point in time, but the results were only mixed at the data interpretation stage—in the overall analysis, the qualitative data carried more weight.

The data was collected in four phases, namely: (a) gameplay observation with notetaking; (b) gameplay metrics were collected by the game during participant engagement with the app; (c) a post-play questionnaire that gave an indication of learning impact, user experience and user satisfaction; and (d) focus group discussions with three different participant groups. In addition, there was also an in-depth interview with a serious game pedagogy expert.

4.1 The thesis commissioner

The thesis commissioner is the Futuristic Interactive Technologies (FIT) research group at the Turku University of Applied Sciences in Finland. The FIT research group is actively involved in exploring cutting edge gaming technologies to apply them in non-gaming contexts. Their current portfolio sees them constructively combining game engine (Unity and Unreal Engine) capabilities and virtual, augmented, mixed and extended reality technologies in the maritime, construction and fire-fighting industries. In each of these industries, FIT undertakes research and development activities that underpin sound teaching and learning principles for creating and testing effective training applications.

Over the past two years, FIT has been particularly active in the fire-fighting industry with a focus on training a wide range of audiences, ranging from school

children to professionals, in both fire safety and firefighting. The fire safety work of the FIT group includes close cooperation with industry experts and end users in testing their applications (Oliva et al. 2019) in order to create virtual environments for multiple technologies (Somerkoski et al 2020) that suit the training needs and pedagogy landscape of the industry. During these activities, FIT have also become proficient in establishing suitable research designs (Tarkkanen et al 2020) for teaching and testing with both VR and AR applications. One of the ongoing studies at FIT, focusing on user experience and usability, saw the development of a VR electric cabin fire simulation (Al-Adawi & Luimula 2019). This application has undergone several iterations and the latest research investigates the inclusion of hand-tracking as a means to improve usability and the overall immersive experience (Luimula et al 2020).

4.2 The experiment application

The training context is fire extinguishing training in virtual reality. The quest in the application is to manage an electrical fire situation. The application is aimed at junior fire-safety trainees as a step between their theoretical lessons and practical training with physical fire extinguishers.

The training scenario (also referred to as the *game* from here on) starts when the trainee enters the gameplay environment. The trainee (also referred to as the *user* from here on) is alone in a corridor with two doors, and some smoke coming through one door opening. The trainee must activate the fire alarm and determine the reason for the smoke to decide further actions. Behind the door is smoke limiting the visibility, but the trainee can see barrels on the right side and two electric power supply boxes, which are the source of the room's fire.

The trainee then exits the room to select a fire extinguisher. There are several options to choose from in the corridor: a fire blanket, a membrane foam extinguisher, a fire hose, and a CO2 extinguisher. The training scenario requires the user to select the appropriate extinguishing method. Only CO2 is valid for electric fire, and the VR app disables all other extinguishing methods in this scenario. When the user picks the extinguisher, she/he must remove the locking

pin before entering the room. To successfully put the fire out, the user needs to stay down, target the spray nozzle directly at the fire and press the fire extinguisher handle. The fire reacts to extinguishing from about two meters distance. The user should stay down, under the smoke, and not go too close to the fire. Once the user has extinguished the fire, they should leave the room and close the door. The user then uses the second door to exit. Figure 5 shows a series of screen captures from various phases in using the application.

Figure 5: View of the tested training applications task and the scoreboard.

The user receives both immediate and summative feedback. Immediate feedback is given in the form of confirming correct selections during gameplay, and scores for engaging in dangerous or incorrect behavior, such as going too close to the fire, are given at the end of play. Summative feedback is derived from the game metrics collected by the app during gameplay.

The game metrics record the time spent extinguishing the fire, whether the fire alarm was triggered, if the pin was removed before entering the room, and that the user closed the door when leaving the room. The game counts these actions as a positive score—the higher the score, the better. The following metrics are scored on a negative scale, choosing the wrong extinguisher, standing in the

smoke too long, or going too close to the fire. The overall scenario score is the sum of all the metrics.

The training scenario does not change and can be repeated as many times as needed. Each learning episode takes about 10 minutes. The feedback is intended to instruct the user how to behave in the initial putting out of the fire.

The application is designed for the HTC Vive Pro virtual reality head-mounted display that lets you walk around in and physically interact with the virtual world. The HMD allows a 360-degree three-dimensional (3D) play area. The eye resolution 1080x1200 pixels per eye and 90 Hz refresh rate offers 110 degrees angle virtual world. These resolution specifications are greatly improved from previous generation headsets and go a long way toward alleviating motion sickness that often limited the use of VR in older devices. The HMD is censored with SteamVR tracking, G-sensor, gyroscope, and proximity sensor. The headset is connected to the computer with a cable. The SteamVR software application platform powers the VR programs.

The virtual interactions are made with HTC motion controllers that utilize catch, release, and teleport functions with one or two controllers in hand. Controllers offer haptic feedback in the form of traditional controller rumbling, which is primarily used to indicate interaction with virtual objects.

The motion tracking works with two base stations that are wall-mounted or installed on tripods to the room. These "lighthouses" are placed diagonally in across from each other. The system works with lasers that identify the HMD's location. The base stations need a clear view of each other in order to successfully track the devices. The traced play area perimeter is set when the system is placed in the room, and a virtual grid warns the user about boundaries of the play area when using the system. The virtual boundary appears when the gamer gets to the play area's limits to prevent the user from physically walking into walls or furniture. The maximum size of the play area is 3,5m x 3,5m.

The training application was created using the Unity game engine. The Unity editor offers 3D graphics, physics simulation, audio, navigation, and more to program the VR application. Unity's XR plug-in is needed to create a virtual reality game. This plug-in manages the target platforms software development kit (SDK).

In this development project, the realistic physical particle effects behavior of the smoke has been specifically made and implemented to enhance the application's physical fidelity.
The application was developed at Turku University of Applied Sciences by the Futuristic Interactive Technologies research group in 2018-19 and various test results have been reported in (Al-Dawi & Luimula 2019).

4.3 The research set up

The research consisted of seven sessions and was conducted in groups of 3 to 8 participants. Each group had participants of a common participant category. Each category presented a different serious game development stakeholder group: fire safety experts and trainers as content specialists, junior professionals (recently trained fire safety) as target users, and final-year game technologies engineering students as game designers and developers. The interview with the pedagogue was conducted to strengthen the insight of serious game instructional design requirements.
The participants first familiarized themselves with VR usage before using the actual application. The VR application collected data during the learning episode while the researchers observed and made notes of the participant behavior during their gameplay. After gameplay, they immediately answered the questionnaire (Annex 1). When a participant group had finished playing and answering the survey, the researcher held a focus group discussion with them. The researcher guided the discussions by asking five questions. The focus group discussions were recorded, and researcher made notes of key topics in the discussions. Figure 6 presents an overview of the experiment design. This process was repeated seven times; five times with expert groups, once with a junior professional group and once with game engineering students.
The data was collected and handled discreetly. All participants got a personal user code which they used when registering to the experiment application and when answering the questionnaire. This research did not collect their names or any other personal details.

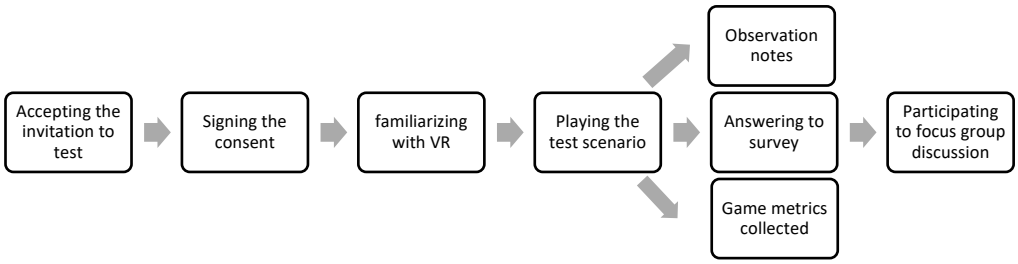


Figure 6: The research set-up of the data collection.

4.4 The study participants

The researchers decided to conduct the experiment with three test group categories to answer the research question. A convenience sampling technique was used to recruit the participants. Researchers recruited the participants by sending open invitations to fire stations, Aboa Mare Maritime Academy, and Turku University of Applied Sciences game technologies engineering students. The criteria were that the application must be previously unseen to the participant and that the participant must be over 18 years old.

The participants received the invitations to take part in this research via electronic broadcasting channels and picked suitable testing times from a schedule. We arranged the testing of all 31 participants into seven groups. Each group had 3-8 persons from the same user category.

The professional test group, who represented the content matter expert stakeholder group, had extensive expertise in the training area. The group consists of 21 professional or voluntary firefighters, and researchers interviewed them in five smaller groups.

The second participant group were second to fourth-year maritime deck officers who have accomplished the STCW (Standards of Training, Certification, and Watchkeeping) Basic firefighting training. Their representative stakeholder group were the experiment application target users.

The third group consists of seven game development students from Turku University of Applied Sciences recruited by their teacher's invitation. These students are fourth-year game technologies engineering students and their competence track consisted of courses covering game design and game development aspects in both entertainment and serious games. Their role was to represent a game designer and developer point of view. All the student participants have extensive gaming experience and are sometimes referred to as experienced gamers, or gamers, in this research.

The fire safety test application had never been played by any of the test participants before this study. The process was described to the participants by means of an informed consent (Annex 2), and all participants agreed with the process before the research started.

4.4.1 The gameplay sessions

The study was carried out in November 2020 over five days at Aboa Mare and Turku University of Applied Science, Turku. The research team consisted of three members, the researcher, the study leader, and the research assistant. When a participant group arrived, they presented the informed consent, and the researcher explained the fire safety application to them.

Participant groups started the gameplay sessions by first familiarizing themselves with VR devices and the physical room environment by playing an unrelated bus inspection application. The research team helped them when needed. When the introductory bus inspection gameplay was over, the participants started to engage with the fire safety application. They received identification codes which they entered into the application to start the gameplay. They played the training scenario for as many rounds as they wanted within the allocated 30-minute gameplay time slot – the number of play-throughs ranged between one and three. The participants were encouraged to verbalize their sentiments during gameplay. The research team observed the participant and took notes of the participants' behavior and talk-out-loud reactions to the application.

4.4.2 The gameplay satisfaction questionnaire

After the gameplay sessions, participants filled a questionnaire that collected information on the gameplay experience and, in this study, it was cross-referenced with the game metrics for analysis. The goal of the questionnaire was to gather first impressions immediately after gameplay, before more details about the experiences were shared in the focus groups. The questions were about learning impact, user experience, and overall satisfaction with the application and the accompanying hardware. In order to establish a link between gameplay and questionnaire responses, the participants used the same identification code in the questionnaire as in gameplay. They returned the filled forms when leaving the VR application room to the focus group discussion. The gameplay questionnaire contained the following constructs and participants were asked to rate each of them on a five-point scale:

Learning impact

- The application taught me something new about fire safety or acting in a fire situation.
- The application helped me to understand how to act in a fire situation easily.
- The application helped me understand how important fire safety issues are.
- I prefer learning using VR rather than other learning methods.

User Experience

- I enjoyed using the application.
- The application was easy to use.
- The virtual environment in the application was realistic.
- I had a sense of being in the application scenes displayed.

User Satisfaction

- I feel my ability to concentrate/focus has improved.
- VR technology has a positive effect on me.
- I would recommend using VR in training courses in the future.
- I feel safe when using this technology.

4.4.3 Game metrics

The VR fire extinguishing application collects data about player actions during gameplay. This study collectively refers to this data as the game (or application) metrics. These metrics are used to evaluate how well each participant performed various tasks within the game. The VR fire safety application collects eight different metrics. The training application tracks and traces players as they select the correct tool and extinguish the fire within the given time limit. The player must press the fire alarm and close the door at the end. Being too close to the fire or standing in the smoke results in a penalty score. The game scores players in both positive (e.g. choosing the right tool) and negative (e.g. being too close to the fire) metrics. The total score presented at the end of the game combines the positive and negative scores. Time-related scores, such as standing in the smoke and being too close to the fire provides an incrementally linear metric starting from zero. Other metric categories include: (a) a sliding scale with capped maximum, such as completing within the time limit that is calculated by subtracting the time taken to complete the fire extinguishing scenario from the maximum allowable time, with a capped maximum score of 50; and (b) all-or-nothing ordinal scale, such as picking the correct extinguisher—the wrong answer gives zero while the right answer gives maximum points.

The complete list of game metrics is given below:

- Time limit score: Max 50, Min 0
- Score penalty for picking the wrong tools: Max 150, Min 0
- Score penalty for standing in the smoke: Max infinite, Min 0
- Chose the right tool: Max 50, Min 0
- Removed the pin: Max 15, Min 0
- Closed the door: Max 20, Min 0
- Score penalty for being too close to the fire: Max infinite, Min 0
- Pressed the fire alarm: Max 50, Min 0

We collected the metrics from each gameplay round and reviewed it per participant category. The data was recorded in terms of whether participants registered a score for a particular metric, using percentages of how many of the

testers scored in each metric. Differences in scoring between tester categories were interpreted with graphs. Also, the progress between the first and last gameplay round was analyzed for participants who completed multiple play-throughs.

4.4.4 The focus group discussions

The focus group discussion followed once the whole group had finished the gameplay session and filled the questionnaire. The discussion was conducted according to group-member preference, either in Finnish or English. All groups discussed the following five questions.

- Would you trust VR as a tool for learning?
- Do you feel confident to use a fire extinguisher after training with this app?
- Have you picked up any personal safety aspects from the app when dealing with fire? Which ones? How?
- Do you feel the app led you to make the right fire-extinguishing actions?
- How would you design and develop such an app?

Each focus group discussion lasted about one hour, and a new question was asked when the discussion of the previous question tailed away. The group was allowed to steer the topic discussion, but the researcher occasionally stepped in to return to the question if the conversation wandered too far from the subject.
To help with gameplay recollection, some of the general observation notes were raised in the discussion—without singling out any participants. The conversation was recorded, and the recordings were transcribed and translated into English. The whole research process is summarized in figure 7.

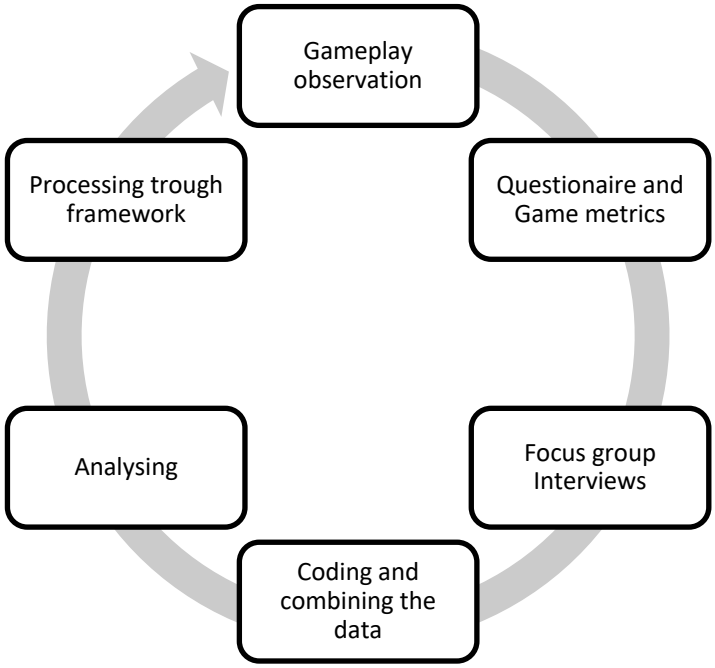


Figure 7: The testing and analyzing process of this research.

Focus group data analysis

The focus group transcriptions were listed in tables, coded, and summarized into categories using the constant comparison method (Boije 2002). The constant comparison method (CCM) ensures all material will be classified and connected to systematize the interview data handling. The method is the core of this study's qualitative analysis, grounded theory approach. Fragmenting the interview data and connecting it again in new ways are the main functions of the method. It was done by coding all relevant themes from the focus group data and framing the themes. Coding the fragments enriched the data and lifted it from the original context of the focus group. This process gave an opportunity to deepen the understanding of the material by reflecting on it within a specific procedure. Conceptualizing the categories helped to find purposeful answers to this thesis' research questions.

Every participant group's codes were allocated to the original questions, summing the grounding weights as a measure of emphasis. To maintain clarity among the

codes and assure a suitable level of reliability, a codebook (Macqueen et al. 1998) was created that explained individual codes that emerged from analyzing the focus group data. The codes were also used a as a basis for defining each codebook theme.

The first groups' coding and grouping into themes were done together with the study leader in order to familiarize the researcher with the required focus group data analysis techniques. Further data coding was done by two researchers with a resulting inter-rater reliability of 80%. The names of the themes were created together.

The focus group analysis was concluded by connecting the different participant groups' input to the serious game development cycle, represented through the Simulation-Games Instructional Systems Design Model (SG-ISD) (Kirkley et al. 2005). To confirm this stakeholder allocation to various serious game development phases, the focus group analysis was also coupled to Serious Game Design Assessment Framework (Mitgutsch & Alvarado 2012).

The quality of the focus group data collection

To ensure that a high-quality focus group data collection and analysis was maintained, the consolidated criteria for reposting qualitative research (COREQ) checklist (Tong et al. 2007) was used. COREQ includes a 32-item checklist for interviews and focus groups. The list presents a compact and comprehensive reporting of the important aspects relating to the research team, study method, context of the study, findings, analysis, and interpretations. Table 2 presents the COREQ checklist items and the respective action taken in this research for each item.

Table 2:COREQ checklist with respective actions from this study

1: Research team and reflexivity	
Personal Characteristics	
1. Interviewer / facilitator	Jenny Lauronen, project manager and researcher in Novia University of Applied sciences.
2. Credentials	B. Eng
3. Occupation	Project manager and researcher
4. Gender	Female
5. Experience and training	M.Eng research methods studies and work experience.
Relationship with participants	
6. Relationship established	No, the researcher and participants didn't know each other before
7. Participant knowledge of the interviewer	Participants knew that the research was part of the researcher's master thesis.
8. Interviewer characteristics	None.
Domain 2: Study design	
Theoretical framework	
9. Methodological orientation and theory	A partially mixed concurrent dominant status method with constant comparison method
Participant selection	
10. Sampling	Convenience sampling. Participants were selected by choosing the required background of the target group and inviting them via their work or study organizations.

11. Method of approach	Via e-mail or spoken word in their work/study community.
12. Sample size	33 were invited and 31 participated
13. Non-participation	Two participants did not come to the experiment.
Setting	
14. Setting of data collection	Data was collected at TUAS and Novia UAS laboratories.
15. Presence of non-participants	There was a research assistant present.
16. Description of sample	We had three participant target groups: Experienced firefighters, trained firefighters, and game developers. They were all over 18 with no previous experience with the study application.
Data collection	
17. Interview guide	The data collection process was described by means of an informed consent, and the focus group discussion was guided with questions.
18. Repeat interviews	No, every participant only took part in the research once.
19. Audio/visual recording	The focus group discussions were recorded.
20. Field notes	Field notes were made during the gameplay observation and focus group discussions.
21. Duration	The gameplay and focus group discussion lasted about 30 and 60 minutes respectively.
22. Data saturation	There were five focus group discussions with experts, and saturation was reached in the third focus group.

23. Transcripts returned	Only the research group reviewed the transcripts with recordings.
Domain 3: analysis and findings	
24. Number of data coders	Two coders coded data.
25. Description of the coding tree	The coding tree was created following the material.
26. Derivation of themes	Themes were derived from data.
27. Software	The data was analyzed in MS Excel.
28. Participant checking	Participants were not invited to review the transcriptions, but they will be presented with an opportunity to read the thesis upon completion.
Reporting	
29. Quotations presented	Quotations are presented, and they can be traced back to participant numbers.
30. Data and findings consistent	Yes.
31. Clarity of major themes	Yes.
32. Clarity of minor themes	Yes.

4.4.5 Pedagogy expert interview

An additional interview with a pedagogy expert was conducted to understand the game from a pedagogy point of view. A senior lecturer from the game development program of Turku University of Applied Sciences volunteered to be interviewed, and the interview took place in January 2021 in Turku.

The interview's goal was to understand the pedagogical perspective of serious game development and find differences in roles and outlooks to the content specialist. The interviewed expert knows the tested application, but the discussion focused on the general pedagogy around serious games. The interview was unstructured and was opened with a general question regarding

how to best use serious games in a training or classroom environment. The conversation flowed naturally, and subsequent questions arose from the expert's experiences as given during the interview. The interview data was collected by means of notetaking.

5 FINDINGS AND RESULTS

During the gameplay, the test participants were observed, and the tested application collected game metrics of their performance. They filled the questionnaire, and they participated in a focus group discussion.
The expert participants were the largest group in this research, and the tests were done in five separate groups. The junior professionals and game developers both had only one focus group discussion.
All expert participants were new to virtual reality. Only four of them said they had used VR before. The attitude was positive, and they said they found participating in this research fun.

5.1 Observation

During the gameplay, the researcher observed the test participant and made notes of particular actions and utterances. Participants were given the opportunity to discuss the observation remarks during the focus group discussions.
At the beginning of the test, all participants had difficulties with the controllers. In particular, all participants found it challenging to perform fine motor movements, such as removing the fire extinguisher pin. They commented that they could not line up the controller with the pin to remove it.
Many of the participants in the expert and junior professional groups were hesitant to open the door that led to the fire. They told in discussions that they were looking for fire gear and fire blankets due to the risk of explosion and poisonous gases. They were analyzing the situation and the equipment available.
The researcher noted that participants in all professional groups had difficulties to navigate the scene. To teleport precisely to the desired location took several tries for twelve of the participants. Participants also bent down to stay under the smoke, but in the discussion, they stated that the scenario did not change as it would in real life. Seventeen participants played a second round. In this second

gameplay instance, the environment and tools were familiar to them, and gameplay took a shorter time.

There was a clear difference in how the professionals and game developers (experienced gamers) approached the situation. The professionals took more time to familiarize themselves with the environment and to analyze the required actions. The game developers, however, did a rapid play-through and memorized the scoreboard's requirements to score better in subsequent attempts. In the focus group, the game developers said they approached the training application as they normally do with a new game. They also mentioned that they were surprised how the fire did not harm them even though they got a penalty score for this. Both groups said they would have preferred more in-game feedback.

5.2 Satisfaction questionnaire

The study participants answered a satisfaction questionnaire that included 12 questions (four questions each) about the learning impact, user experience, and user satisfaction (§ 3.3.2) immediately after finishing the gameplay rounds. The answers were given in Likert-scale 5 (strongly agree) to 1 (strongly disagree). For the graphical results reporting, the 5-point Likert scale was retained (Figures 8 and 9), but during the analysis phase, it was decided to reduce the Likert scale. Answer scores of 4 and 5 were viewed as agreement, while answer scores of 1 and 2 were seen as disagreement. A score of 3 was disregarded as neither agree nor disagree. This reduction was made because the number of participants in the junior professional and game development groups were too low to make sensible assertions with a 5-point Likert scale.

In addition to reducing the Likert scale, the junior professional group responses (3 responses) were assimilated into that of the professional group. Both of these groups were familiar with firefighting, and they responded from the same perspective. The game developer group returned 6 questionnaires. The graphs in Figures 8 and 9 show the distribution of responses to each question and the percentage of participants per rating given. The graph in Figure 10 presents how each group rated the different questionnaire categories.

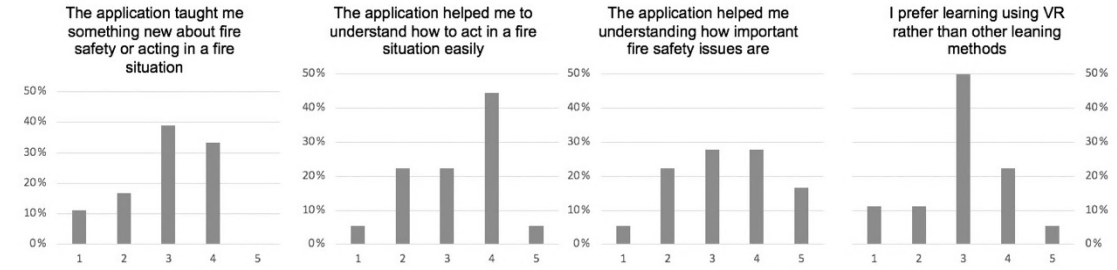
Professional firefighters and junior professionals

The expert's answers for learning impact were widely spread. All four questions had similar magnitudes of agreement and disagreement.

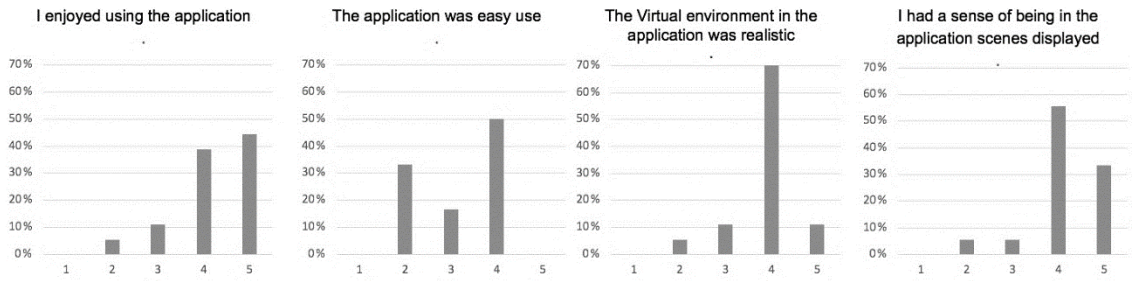
The overall user experience answers were positive in the expert group. Over 80% of the experts enjoyed using the application (question 1), and 80% also felt immersed (question 4). The ease of use question got 30% disagreement and 50% agreement.

The questionnaire got the best feedback in the area of User satisfaction from the experts—almost no disagreement in any of the questions. Over 90% felt safe when using VR technology and almost 90% of the experts would recommend VR as a training environment for fire safety.

Learning impact



User experience



User Satisfaction

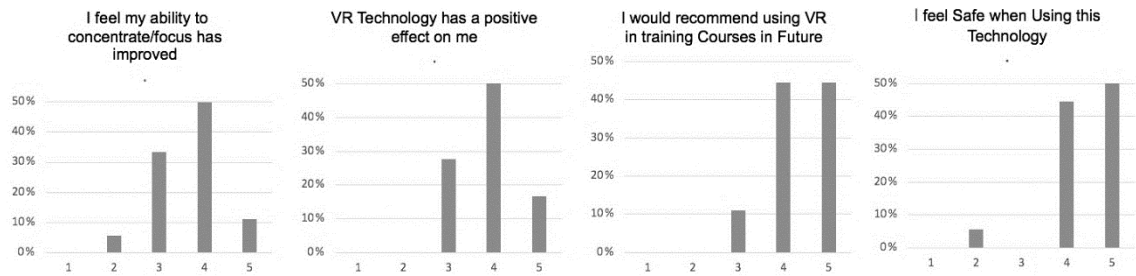


Figure 8: Questionnaire results from combined experts and junior professionals.

Game developers

Game developers preferred VR over the other learning methods, which was the most preferred proposition on the learning impact set of questions. In other questions, the response opinions were spread (Figure 9). For the user experience section, the enjoyment of the usage, and the application's realism got about 80% positive feedback. The responses about ease of use were evenly divided. The user satisfaction group section got the highest rates, especially "I would recommend using VR in training Courses in Future" and "VR Technology has a positive effect on me" got over 80% positive response.

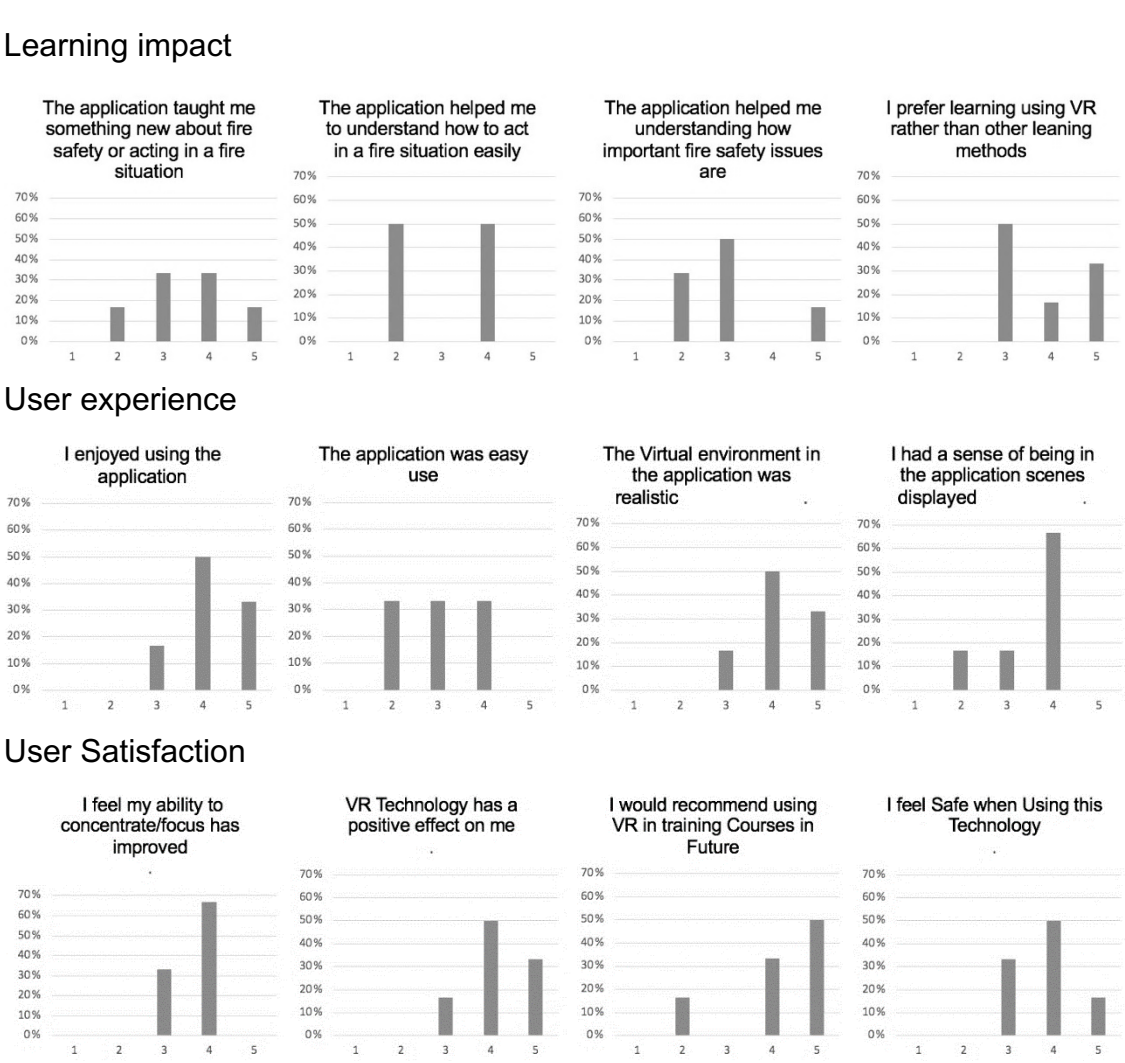


Figure 9: Questionnaire results from experienced game developers group.

When comparing the graphs, appeared that the difference between the participant groups were minimal. The trend across the three categories was the same, even though the participant groups had such different backgrounds. The answers looked similar across both groups. Both the experts and the game developers gave the highest number for the questions about user satisfaction and the lowest scores for learning impact. (Fig.10)

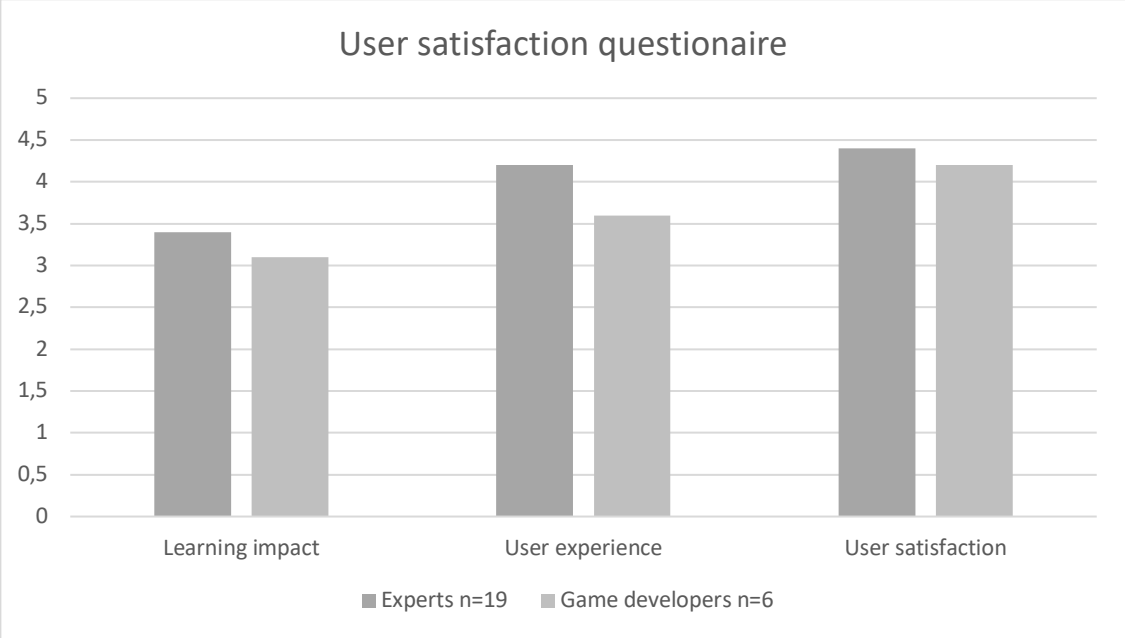


Figure 10: Questionnaire results from experts and game developers.

5.3 Game analytics

The game analytics collected the gameplay metrics from eight different in-game activities, and the total score was a summary of all metrics combined. However, one of the metrics (picking the right tool) is not shown, as it was the only device enabled, and all testers scored in that. The collected data from 21 professional firefighters and three junior professionals were combined due to the small number of junior professionals and the two groups' parallel expertise. Seventeen of the participants played more than one round of the game. The game developer group has data from seven participants. Unfortunately, the game developer data is only from their last play-through — the application did not save earlier play sessions.

This data loss might have been due to their quick play-through strategy of learning from mistakes and starting over before the actual end of the play-through when metrics are stored for analysis.

The scores as recorded by the game were widespread (§ 3.3.3) with values varying greatly between 0-400 in both positive and negative axes, and many of the scores were zero. This made that the values did not lend themselves to coherent visualization. It was decided to rather analyze the proportion of participants who scored in each metric, and to present the results as percentages (Figure 11). As an example, one would interpret the last set of bars in Figure 11 as follows: between 70% and 80% of professional participants pressed the fire alarm while less than 20% of the game developer group did so.

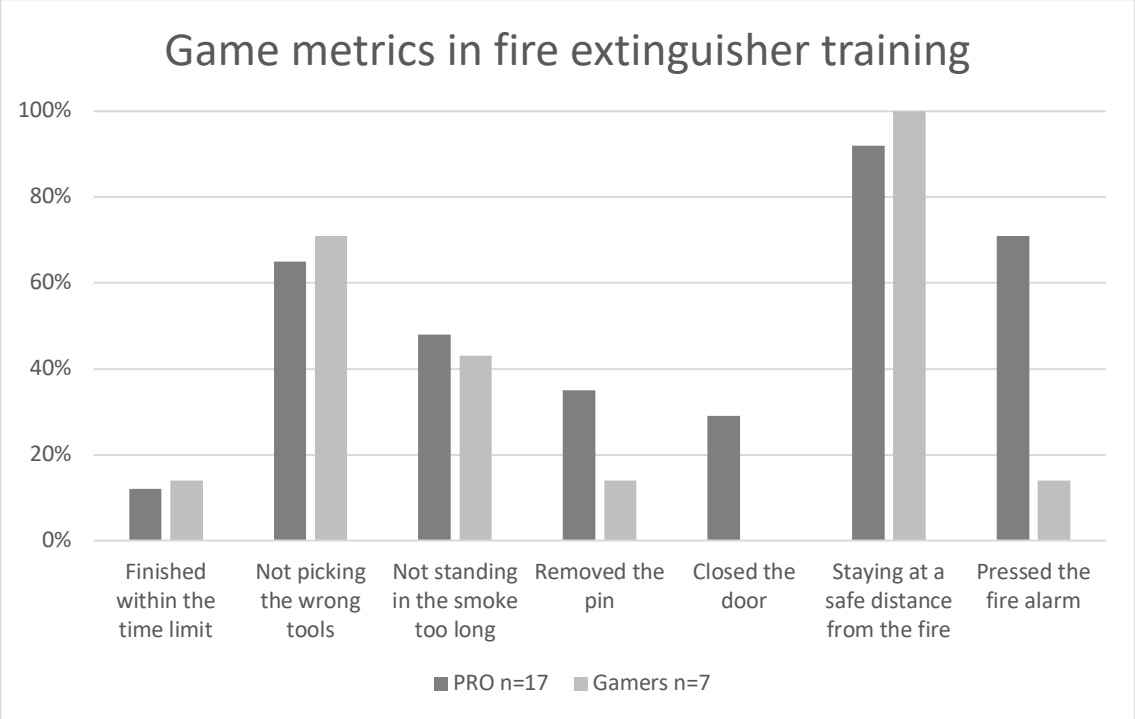


Figure 11: Proportion of participants who scored in-game metric in their last round.

The progress of the experts over the rounds, from the first to last gameplay round, is shown in Figure 12. The first bar of each comparative pair of bars shows the proportion (in percentage) of participants who managed to score in their first playthrough, and the second bar shows the proportion of participants who

managed to score in their last playthrough—this analysis was done only with those who registered multiple playthroughs.

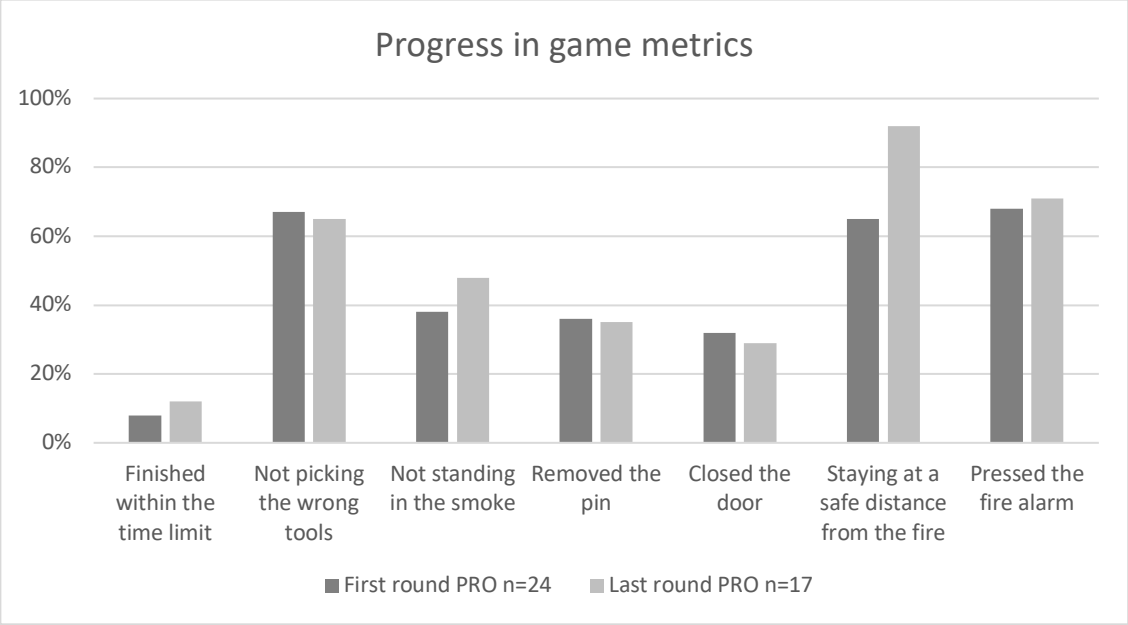


Figure 12: Comparing the proportion of scores recorded in the first and last playthroughs

5.3.1 Game analytics findings

The time score (measuring the time it takes to extinguish the fire) was the most challenging, and only about 10% of the participants scored in it. There was a slight improvement in the number of participants who managed to complete the fire extinguishing task over the rounds.

In the score of picking the tool, both game developers and professionals scored. In the focus group discussions, the professional firefighters explained they also wanted to have the fire hose ready when they opened the door, hence, they continued trying that tool. The game developer play-through strategy led them to select different tools in each play-through. There was no score difference between the first and last rounds for the expert group.

The standing in the smoke metric progressed from first to last round. During gameplay, the researcher observed participants bending down, and in the discussion, the participants stated that the visuals did not lower as they went down. The discussion gave the impression that the calibration for bending was set too low, and it was not aligned with the virtually visible smoke. This created a challenging situation to stay under the smoke in the limited testing space.

When using the extinguisher, the users were unaware that they must remove the pin before entering the room. Participants found it challenging to remove the pin but were able to perform the task with some assistance and extinguish the fire. However, many of them only removed the pin inside the room where the fire was and did not score.

According to the metrics, closing the door was missed by many. The score did not improve in later rounds, and no game developers scored in this task.

The proportion of expert participants who learnt to stay in the safe distance from the fire raised from 70% to 90% over the rounds. Everybody in the game developer's group was able to score in that. In the discussion, the experts said that teleporting made it difficult to get to the proper distance to extinguish the fire. If they teleported too close, it meant they had to turn around to go to a safe distance and their expert training discourages turning their backs on the fire. Three of the game developers said they jumped to the fire first, but according to the metrics, they found a better distance.

Pressing the fire alarm was the positive metric, and most of the experts did that. In the discussions, the experts commented they did not get feedback whether the pressing succeeded. The lack of feedback when pressing the fire alarm might have caused a reduction in the score from tried actions to press the alarm.

5.4 Interview with the virtual learning pedagogy expert

The interview of the learning pedagogy expert highlighted the role serious games could have in a learning process. During the interview, the pedagogy expert explained his approach to determining a suitable design for a serious game.

The instructional design of a virtual learning environment starts from questioning the bigger picture of the game—is it meant to supplement a current learning program or is it a stand-alone game. When designing the game's usage as a part of learning material, the game's role in teaching the topic must be clarified so that the game's learning episode can add value to the existing learning outcomes.

When the game's role is clear, the story, challenges, and reward system must be designed within the context of the subject matter to ensure an effective learning impact. Serious games can be used in several ways for testing and assessment. Among more creative variations, games can: (a) use reward mechanics that the player sees as they progress; (b) have explicit sets of questions post-gameplay; or (c) collect stealth metrics with a summary of achievements presented at the end of a gameplay round. No matter which assessment strategy is implemented, the metrics must always be in line with the learning goals.

In summary, serious game design from a pedagogy point-of-view must happen on two levels: (a) the elements in the environment around the game (e.g., the course, physical location, target group); and the game itself (e.g., does the game address the outcomes appropriately).

5.5 Focus group discussion

The focus group findings are summarized as codebooks in a series of tables. The transcriptions were analyzed using the constant comparison method and the codes emerged from the data without using any prior codes. Similar codes were consolidated into data representative themes. In this section, the summary of the codes and themes are organized according to participant group.

Each participant group's data is presented in separate codebook that have identical structures. However, the codes vary between the groups. The codebooks consist of three columns that name, define and provide example statements for each code. The examples of the phrases said during the interview have been included to explain how each code was formed. The codebooks also show the themes and the codes that make up each of the themes. These themes

are used later on in defining the roles of the different game development stakeholders during the game development process.

The last table in this section presents the overall groundedness of each code and theme across all three participant groups. Groundedness refers to the code frequency, or how many times each a code was mentioned.

Codebook for expert firefighters

The codebook of the experts presents the topics that the experts spoke about. The realism of the scenario and the learning approach were the major topics in their focus group discussions. Especially the discussion about functional fidelity was rich.

Table 3: Codebook for experts

Experts		
Code	Definition	Examples
Theme: Realism of the scenario		
Functional fidelity	Simulation accuracy. The realism of the game scenario during the learning episode.	"Does not present the heaviness of the work," "Necessary fire gear and gas mask are missing."
Physical fidelity	Refers to interaction and feedback (visual, sound, haptic, etc.) that the user is getting or requiring during the learning episode.	"There should be sound - it's always noisy," "Moving under the smoke is important, and it did not visualize here."
Theme: Teaching approach		
Purpose	Purpose and benefits of virtual training.	"supports training," "Could work as test what is learned."
Learning feedback	Feedback given to support learning during and after gameplay.	"Immediate feedback for dangers is important" "Doing things in the correct order is essential, and that's not counted in score now."
Learning options	Choices offered during the learning episode.	"Other types of extinguishers should be presented- and what happens if you pick it."
Learning outcome	Learning results that the game could or should generate.	"Can teach decision making alone," "Must teach right performance."
Place to use	Suggestions of places for the use of such VR-training.	"VR-training can be carried out in a small place."
Target learner	Comments about whom the learning episode would fit.	"For fireman, training scenario needs to be more detailed," "Electrical fire is relevant for electricians."

Learning potential	Comments relating to the role of VR -training and its potential.	"Relevant as the method," "Does not replace real training."
When to use	Comments about when to use the VR training along the learning path.	"Could work as a test when there are right goals and tasks."
Theme: Usability		
Interaction efficiency	How interaction supports the use of the application.	"I wish to have smoother ways of moving" "I pressed first the alarm - did it activate?"
Entertainment value	Comments related to the entertainment side of training.	"Fun, immersive," "It was fun."
Technology flaw	Comments related to VR device usability	"Having difficulties with controls," "Experienced technical difficulties in VR."

5.5.1 Codebook for junior professionals

The junior professionals paid attention to fidelity and the teaching approach, especially the adequacy of the training. This group was interested in the same topics as experts, but they did not talk about target learners and learning potential.

Table 4: Codebook for junior professionals

Junior professionals		
Code	Definition	Examples
Theme: Realism		
Functional fidelity	Simulation accuracy, the realism of the game scenario during the learning episode.	"The dangers and the physicality of this work do not appear here," "Dangers like poisonous smoke and explosion are not taken into consideration."
Physical fidelity	Refers to interaction and feedback (visual, sound, haptic, etc.) that the user is getting or requiring during the learning episode.	"Thinking of the ways how to handle the extinguish," "Here I feel safe going close to the fire, in a real situation not."
Theme: Teaching approach		
Purpose	Purpose and benefits of virtual training.	"This allows to learn in peace, without rush, thinking first and then acting," "Lowers the threshold and eases the fear."
Learning feedback	The feedback given to support learning during and after gameplay.	"The fact that you can burn yourself if you go too close does not appear here."
Learning potential	Comments related to the role and potential of VR -training.	"Does not replace real training," "When there is fire, you need clear instruction."
Learning outcome	Learning results that the game could or should generate.	"Realized that there is fire and I must locate the extinguisher," "Brings confidence to behave right in a fire situation."
When to use	Comments about when to use the VR training along the learning path.	"Supports training," "Works for the introduction."
Theme: Usability		
Technology flaw	Comments related to VR device usability.	"Experienced technical difficulties in VR."

5.5.2 Codebook for game developers

The focus group discussion with the game developers highlighted the feedback and learnability. This creates a new theme: playability, which played a significant role for game developers. They were also interested in physical fidelity, a representation of authenticity within the game. In the discussion, this group highlighted their trust in learning with virtual reality.

Table 5: Codebook for game developers

Game developers		
Code	Definition	Examples
Theme: Sense of realism		
Functional fidelity	Simulation accuracy, the realism of the game scenario during the learning episode.	”I jumped straight to the fire and did not die,” “Details like pin add realism.”
Physical fidelity	Interaction and feedback (visual, sound, haptic, etc.) that the user is getting or requiring during the learning episode.	”audio to help you and giving feedback,” “No alarm of the fact that you can burn yourself if you go too close.”
Theme: Playability		
Entertainment value	The entertainment side of training.	”I took it as a game,” “Did not like the pin; it would be better just spray.”
Immediate feedback	The feedback that supports learning as the episode progresses.	”haptic feedback or visualization could appear when you are in danger,” “I need to get some feedback of the danger.”
Post-game feedback	How the reward mechanic can be used to explain the correct actions in the game.	”When practicing as in the games, you see results only after and learn what you should have done,” “Prefer to

		have feedback after the round"
Learnability	Comments on learning how to use the game.	" During the first round, I didn't know what was expected," "I did right things after one round."
Theme: Teaching approach		
Instructions	Comments about instructions and guidelines.	"It (training episode) presents a different kind of extinguisher but does not explain why to pick that one."
Learning outcome	Learning results that the game could or should generate.	"It's good to see there is so much to do in the actual situation" "Teaches details like close the door, press alarm - that gives an understanding of the realism."
Learning potential	Comments related to the role and potential of VR training.	"You must practice this in real life."
When to use	Comments about when to use the VR training along the learning path.	"I've used it (extinguisher) before, and this was the reminder of how to choose the right one."
Theme: Tech acceptance		
Tech acceptance	Acceptance and potential of VR technology as a learning platform.	"Relevant as a method," Yes, trust 100%".

5.5.3 Weights of emphasis across the themes

When the focus group interview material was coded, the groundedness (frequency) of the subject matter was also recorded (Table 6). Not only were the themes different between participant groups, but the code groundedness also differed between groups.

The functional fidelity got the most considerable interest within expert groups—they mentioned it several times in every discussion topic. As functional fidelity was the most analyzed code, the second biggest was the learning outcomes. These two areas were common for both content specialist groups. For experts, however, the pedagogical questions played an essential role, shown in codes such as learning feedback; purpose and learning potential; and target learner. Since junior professionals were closer to being target users, they talked a lot about physical fidelity that presents the actual doing. Also, they discussed the learning potential, purpose, and learning feedback. Physical fidelity and learning feedback together present the experience authenticity.

For game developers (game development students), the playability came to prominence. Other participant groups did not pay much attention or did not verbalize this. The playability theme included correctly timed feedback and the learnability of the game. However, fidelity and learning outcome played an important role in discussion with them as well. Gamers were positive about virtual reality, and they specifically mentioned they would rely on the method.

Table 6: Summary of the weights of emphasis across the themes

Code	Experts Firefighters (n=21)	Junior professionals (n=3)	Game developers (n=7)
Theme: Realism (groundedness = 200)			
Functional fidelity	121	22	8
Physical fidelity	23	12	14
Theme: Teaching approach (groundedness = 210)			
Purpose	21	5	
Learning feedback	30	5	
Learning options	7		
Learning outcome	62	6	16
Place to use	1		
Target learner	21		
Learning potential	21	7	2
When to use	2	2	1
Instructions			1
Theme: Usability (groundedness = 22)			
Interaction efficiency	11	1	
Entertainment value	3		
Technology flaw	5	2	
Theme: Playability (groundedness = 34)			
Entertainment value			7
Immediate feedback			10
Post game feedback			7
Learnability			10
Theme: Tech acceptance (groundedness = 4)			
Tech acceptance			4

6 CONSTANT COMPARISON ANALYSIS

Analyzing the codes retrieved from focus group discussion data with a constant comparison method means connecting the themes to the research questions through the selected frameworks and models. The main functions of the method are to first fragment the discussion transcriptions into their smallest parts (codes) and connecting them again in new ways. The data from the focus group discussion, (summarized in tables 2-5) was analyzed for each group and compared with each other.

In this research, once the thematic network was established, the analysis continued by comparing the interview data with the System Design Model (§ 2.2.1) and the Serious Game Design Assessment Framework (§ 2.2.5).

6.1 Expert group's focus of interest

The expert participant group's primary interest was in the realism of the scenario. They talked about their experience and how it simulated the reality they are familiar with. This theme had the most extensive interest for this group. They were also thinking about the purpose of the training scenario, the target learner, and learning goals. They paid attention to the relevance of the scenario, thinking to whom it could be targeted and which learning goals would be relevant to each target group. The right behaviors and safety perspectives were paid attention to in every question, even if it was not asked. They saw the bigger picture and the consequences of each element in the learning episode.

Some participants of this group are trainers in this area. They also discussed how the training fits the learning path and which things in that path would be most helpful to practice in virtual reality. In their codebook, the code "target learner" distinguishes what learning goals fit each group and how they can be implemented to their training.

6.2 Junior professionals focus of interest

The trained junior professionals discussed the training episode's fidelity and teaching approach. As they had recently participated in fire safety training, they talked about their experience and compared the training application's functionality with that. In the teaching approach, they highlighted the purpose, learning outcome, and interaction efficiency.

The junior professional group noticed the elements in realism and learning goals that weren't in place in the scenario, and it was also noted during the gameplay observation. The junior professionals relied more on their recently completed training material and made decisions with less questioning of the scenario. The group commented on their experiences in similar areas to professional groups, but they could not verbalize their thoughts as profoundly as the expert group. For example, when junior professionals were thinking about their actions' safety, the experienced professionals did the same and listed reasons of what could have happened and what they would have needed to do to be safe. The experts also indicated that only a team of professional firefighters with specific equipment would be able to handle such a situation—also something target users were not able to express.

6.3 Game developer group’s focus of interest

The game developer group commented mostly on playability and fidelity. Much of their interest in playability was subject to the interaction and feedback. They ideated how to build reward systems and how to offer immediate feedback for the user. They also had many ideas for interaction, game mechanics, and how to help the users score better. However, the thoughts were not necessarily synergistic with the training goals. Since the training application was aimed at professional training, the design requires understanding the situations and previous education or instruction. The need for instructions was interesting for them, and they discussed appropriate feedback timing during many of the questions posed in their focus group discussion.

The game developers had parallel ideas with professional groups about usability and developing physical fidelity. The trust and expectations for virtual reality as a learning media were high in this group, and they also discussed the entertainment value of the training episode.

6.4 Comparison between the participant groups

When the professionals were discussing fidelity, they focused on the broader aspect of the scenario. All themes and the majority of the codes are the same between experts and junior professionals in this research. However, the topic perspective was slightly different. The experts spoke of the training situation's holistic experience and concept, while the junior professionals were more inclined to talk about the training scenario's task level. As the situation and its goals were new to the game developer group, they interpreted fidelity as functionality.
In functional fidelity and training, outcome perspectives vary clearly (see table 7). Some of the game developers' ideas were contradictory to the expert's thoughts in this area. The interaction and feedback themes vary, as their expertise took the game developers to discuss different aspects of playability. For gamers, this is also a matter of deeper interest. They brought solutions and suggestions on how to support the learning episode with reward mechanics, while experts approached it more from a learning point of view. Table 7 illustrates the different focus areas for each participant group by highlighting some of each groups' key comments.

Table 7: Comparison of the key elements across the test groups.

Experts	**Junior professionals**	**Game developers**
Functional fidelity		
“Must be clear with goals and actions.”	“Dangers like poisonous smoke and explosion are not taken into consideration.”	“Details like pin add realism.”
Learning outcome		
“Must teach right performance.”	“Brings confidence to behave right in a fire situation.”	”It's good to see there is so much to do in an actual situation.”
Codes found in different categories: teaching approach (professionals) and playability (game developers)		
Learning feedback	Learning feedback	Immediate feedback
“Immediate feedback for dangers is important.”	“The fact that you can burn yourself if you go too close does not appear here.”	”Haptic feedback or visualization could appear when you are in danger.”

6.5 Analyzing the results against the game development cycle

When reflecting on the data analysis results in the context of the game development cycle, each stakeholder's roles can be set in their places. The need for context understanding clears when answering the question "by whom should we test professional training applications and who should be involved in each stage of the game development cycle?". We must consider that expert resources are most limited, which makes their involvement in the correct phases especially critical.

This research investigated four different stakeholder groups as contributors to developing successful training apps: (a) the experts represented the content specialist; (b) the junior professional group represented the target users; (c) the group of final-year game development engineering students represented the development team and any testing that can be done in-house when developing

games; and (d) the pedagogy expert was interviewed, and although this data was not coded, it was included in the analysis.

The participant groups were matched to the game development cycle through the Instructional System Design Model by Kirkley (2005). The phases of the model are presented in section 2.3, The game development cycle. Table 8 connects the development cycle to this study's data analysis by linking the codes that best fit each phase of the development cycle. In this way, the participant group who made the most significant contribution to those codes are linked to the closest matching phase in the cycle.

Table 8: Analysis of the involved stakeholders in the game development cycle.

Development cycle phase	**Linked codes from focus groups and interview**	**Participant groups most responsible for the codes**
Analysis: Instructional theory Needs analysis and target audience External data	Learning options Place to use Time to use Target learner Learning potential	Pedagogue Experts
Concept: Learning methodology Game features	Learning outcome Functional fidelity Learning options Purpose	Pedagogue Experts Game designers
Design: Character design Design lessons Design media Storyboards Assessment design	Physical Fidelity Learning feedback Learning outcome Functional fidelity Learnability	Game designers To evaluate storyboards and in assessment design, experts and/or pedagogue are needed (content-specific)
Q&A + prototype: Bug testing Usability testing Play/fun testing Learning testing	Functional fidelity Physical fidelity Purpose Playability Learning feedback Learning outcome Interaction efficiency Technology flaw Entertainment value	Junior professionals or target users Gamers Game designers
Implementation: Interactive lesson plan Learner material / Game	Time to use Purpose Learning potential Place to use Instructions	Pedagogues or experts Game designers Target users
Summative evaluation: Test for instructional quality Needs assessments	Time to use Purpose Learning potential Place to use Instructions	Pedagogues and target users

6.6 The Serious Game Assessment Framework analysis

The themes extracted from the focus group data, together with the analysis against the game development cycle, were compared with the Serious Game Assessment Framework's (SGAF) elements. The goal of this analysis was to uncover the importance of correct timing for involving domain-specific expertise in relation to the game's overall usefulness. The assessment framework is used here as a design tool to address the necessary decisions and definitions required for successful serious game design. This method also allows the analysis of the game system's coherence and cohesiveness of the elements in relation to each other, and to the game's purpose. The SGAF elements are introduced in section 2.3.

The summary below (Table 9) demonstrates the importance of expert involvement in the analysis and concept phases. Learning goals and target groups are also prevalent in this assessment. All elements of this analysis were connected to fidelity at some level.

Table 9: Summary of SGAF in relation to the game development cycle and participant groups.

SGAF's category	**Decision phase of the game development cycle**	**Categories from the codebooks**	**Participant group who should be involved**
Purpose	The purpose is determined in analysis and concept face and evaluated in Implementation.	The realism of the scenario Teaching approach highlighting target user	Only experts brought input in this area.
Content & Information	This handles with fidelity and take place mainly in the concept phase, and is tested in the prototype.	Realism Teaching approach	Only experts can help in concept determination. Once it's throughout done, the target users can do the testing – but if something is missing, they won't be able to verbalize that.

Game mechanics	Definition of game mechanics is done in the concept phase and tested in the prototype	Teaching approach, mostly feedback Realism Playability	Only experts can help in concept determination. Once it's throughout done, the target users can do the testing – but if something is missing, they won't be able to verbalize that.
Fiction & Narrative	Definition of narrative happens in the concept phase. The user experience and usability testing are evaluated Q&A / prototyping.	Realism Teaching approach, especially purpose	Only experts can help in concept determination. Target users can do the testing and suggest the improvements – but if something is missing, they might be able to verbalize that.
Aesthetics & Graphics	Aesthetics and graphics are primarily determined in the concept and design phase.	The realism of the scenario Realism Playability	The target users seem to be able to test this in the prototype. In later stages, the regular users can the usability and user experience if the bases are correct.
Framing	Framing the game is determined in the analysis and concept phase.	The realism of the scenario Teaching approach	Experts must do this.
Coherence and cohesiveness	This must be done in summative evaluation.	The realism of the scenario Teaching approach	This must be done by experts. Possibly in the concept phase, the coherence and cohesiveness can be evaluated through documenting. It would be a helpful tool to validate the concept in a formal way.

7 DISCUSSION

This study aimed to determine the applicability and timing of the different stakeholders in designing and developing a VR training application or serious game. We assumed the participants would primarily focus on user experience and usability issues. During the study, especially in focus group discussions, it became clear that the expert is needed at the very beginning of the game concepting. Their involvement beyond the prototyping phase would be too late. They are interested in content specific aspects, such as learning outcomes, the target learner and realism. If these points are created with too little understanding of the context, provided by the experts, the application runs a real risk of failure.

These findings mainly follow Olsen's (2011) ideas, where professional testers are involved in the ideation (refers to analyzing phase in ISDM) and concept phases. Olsen also sees the need to interview target users in the pre-development phase. If the game's target group has limitations, for example, disabilities, it may influence interaction with the game. They also encourage developing a playable model of the storyboard and testing it with the target group. They also urge to interview test groups, as well as have them answer questionnaires. During this research I realized why Olsen insists on a mixed-method approach; our questionnaire did not uncover differences between the participant groups. The questions were not at the right level for this stage of application development. In the interview, the questionnaire responses got deeper explanations.

From a game testing informativeness point of view, the questionnaire results were interesting. The survey gives the impression that participant groups experienced the application similarly. This is concerning from the reliability aspect of testing because other study outputs (e.g., the focus group discussions) showed apparent differences between the groups. Study results from short questionnaires should, therefore, not be analyzed in isolation, but rather be combined with qualitative data for a richer understanding. However, questionnaires in various forms are still widely used and relied on for testing user experience and usability. A suitable survey could have described the development progress, but it did not offer an answer to the study hypothesis in this case. In this research, we learned that

testing a questionnaire would ensure it brings useful information. Furthermore, when using questionnaires, the participants must be interviewed and preferably observed to comprehensively understand the answers behind the lines.

The pedagogy and implementing the learning goals was recognized as a challenging task in the literature review. In the pedagogy expert's experience in the interview, the same phenomenon was highlighted. The serious game's role in the learning path and learning goals should be the starting point of the concept. Only then, can balancing instructional and game design happen. Pedagogical know-how was recognized in instructional-oriented models and frameworks (Braad 2016), but literature remains unclear about when, in the development cycle, to actually seek and action pedagogy expertise. Instead, pedagogues are largely expected to participate throughout the design and development cycles of serious game development. In most, if not all, cases this is not possible because such experts are not employed to oversee serious game development. This study shows that the decisions where pedagogues are most valuable lie in the research and concept phase of the serious game development cycle.

The junior professional group’s ability to recall their prior training experiences, which was suggested by Alexander et al. (2005), came out clearly. The psychological fidelity resonates with their user experience and gameplay behavior. It is the same phycological response instilled in VR training as it would be in a real-world situation. During their gameplay, the scary situation of fire in the closed room was obvious when observing them practicing. In the focus group, they spoke about their training experiences how it was daunting. They said it was good to have time to think in one’s own space and make decisions independently—something that is not possible in real-world training. Own pace in practicing can help to manage the fear, they said.

The dangerous situations that could arise in the scenario, like explosive space and oxygen superseding the CO-extinguisher in a closed space were discussed. The junior group felt that this VR scenario would be excellent for experienced firefighters, provided that the learning goals are clearly aligned with the scenario. The message that can be understood from this, is that it would be fine to abstract, or miss some real-world elements in the VR case, as long as the learning goals

are clearly communicated beforehand. Target groups and clear learning goals are inevitable to avoid discrepancy.

The SGAF as a design tool, offering coherency of all game elements, is a substantial help in serious game design, even if it's meant for assessment. This was done because there is a lack of serious game design models or frameworks that give practical support in combining both instructional and game design (Iuppa &Borst 2010, Cowan & Karpalos 2017, Fowler 2015, Mitgutsch & Alvarado 2012, Naranjo et al. 2020,Mayes & Fowler 1999, Alvarez & Djaouti 2010, Avila-Pesantez et al. 2017, Fullerton 2014,Olssen et al. 2011). This was also noticed by Braad et al. (2016) in their work of finding useful design frameworks. Participation of the different stakeholders was not specified in any of them, and it seems to be a needed topic requiring further research. As content specialists are the rarest stakeholder group for testing, there is no reason to wait until the application is ready, but rather to use their time in the very first ideation and concepting meetings. Some literature names this phase, "research" and also the stakeholders as "researchers." Although such a generalized set of researchers may be workable in some cases, skills acquisition VR applications need a stronger specification that these early phases are the most impactful timing for the use of content specialists.

The game metrics and feedback pointed out many interesting questions. However, we did not research it throughout because the study setup was not designed for this specifically and it was outside the scope of the thesis. Although the experts were not experts in virtual reality, they scored higher in many areas than game developers. The reason is that the scores measured actions that were familiar to them, and they maintained their performance accordingly. Importantly though, the game metrics appeared to measure what experts considered as the important actions during gameplay. Over the rounds, some metrics did not quite progress, and in focus groups, two reasons were pointed out. Firstly, the limitations of the right actions caused negative scores continuously. In other words, the experts were expecting other, more viable, options to be available in the scenario. This expectation discrepancy could once more have been alleviated if the learning goals were more clearly communicated. Game metrics have a

cross-referencing function with post-play debriefing and should always be in close relation to learning goals. The second challenge was the feedback. Some scores, for example, removing the extinguisher's pin, did not progress because the information when it must be removed was not available. This would have been particularly useful learning information for gamers, who extensively looked to improve their scores by using information presented on the scoreboard before returning to the quest. Another feedback issue was raised by experts for learning purposes. In their opinion, serious games should not teach dangerous habits, and a lack of feedback against wrong and dangerous behaviors do not instill safe real-life practices. The third form of missing feedback is a technical flaw that happened with the fire alarm, which did not activate when touched. Experts also mentioned that they would have appreciated some indication (sound or lighting) in VR that the fire alarm had been activated. Game developers, on the other hand, were satisfied in learning from the post-play metrics that they needed to trigger the fire alarm. Once game developers were aware of this required action, they also noted missing the immediate feedback when triggering the alarm. This once more shows that the experts knew the alarm needed to be activated and that there should have been an audial or visual cue for this—information that would have been valuable early in the game ideation.

The right scales and balances for the metrics and feedback on how to score is an essential part of making a serious game impactful. Game metrics (analytics) should not be purely viewed as a scoring system but can become an essential game design tool when applying an iterative development process. Unreliable or incorrect measurements could point to usability, feedback, learning outcome and other design aspects that could be addressed in subsequent game development cycles. The motivation, progress, and fun elements that gameplay brings to learning can be tested early with paper models or rough prototypes with experts to get all the required info toward appropriate game mechanics.

8 CONCLUSION

The material revealed that the experts have to be involved in the early stages of the design process. In light of the data, the scenario's realism and training impact are the experts' primary focus. The game designers need to be able to produce an appropriate description in order to create solutions with a desired outcome.

This research shows that usability and interaction are recognized in similar ways across all participant groups. In the later stages of testing, involving the experts is too late. If fidelity and the teaching approach is determined correctly in the analysis and concept phase of serious game development, target users can be utilized to test the user experience, and even regular (non-target) users can be used for usability testing.

For ideation and conceptualizing the training application, this research proposes to use the experts. The target users do not have the necessary ability to recognize or communicate the flaws in realism and teaching approaches. They cannot be used in learning goals creation as they do not know what must be mastered, much less how to train such mastery. Pedagogues also need to be consulted in these early phases to ensure a significant learning impact and that the game's role in the learning path benefits target learners.

Planning and drafting alone, however, do not equate to successful implementation. Therefore, early storyboard or prototype testing should involve experts to make sure the designs and development have not deviated from earlier expert input. This does not require deep consultations, but short demonstrations or desktop playtest sessions with expert feedback would be sufficient. For example, the experts can comment on how the scenario works and what prevents them from acting correctly. Also, the junior professionals reacted to the same phenomenon, but they did not question the training scenario when it limits the correct behavior. In other words, they might not help the game design team get the game's premises right.

The game developers' strength in testing is their focus on playability and feedback. Both of these significantly influence the learning impact. Game developers are possibly more free to ideate the gameplay without the constant

distraction to replicate a realistic environment and set of tasks. Their role could find a place in creating entertainment over a solid knowledge foundation from experts and pedagogues, bringing about new and innovative ways of teaching the right skills and actions.

The game design team needs a straightforward process with an understanding of wholesome serious games. There are many options. Choosing any of the existing models will give benefits - System Design Model by Kirkley, Serious Games Usability Testing by Olssen, or one of the models presented by Braad in Processes and Models for Serious Game Design and Development. The most suitable model is case-specific. The models should ensure that the phases of testing cover the required information in each stage. The use of a model helps to avoid wasting resources when developing in the wrong direction.

If testing is done with the wrong stakeholders, the results are not valid for the game's goals. The application is not effective in teaching a topic because of poor fit to the purpose or incoherence within the subject. Loss of focus causes the content to become too rich and complex and reduces learnability and, more importantly, playability. On the counter side, oversimplifying the quest produces the wrong confidence effect or game that cannot transmit the desired learning results.

The good news is that a small group of participants can do the job. Testing can reach sufficient saturation with even a group of five participants in human interaction usability and user experience testing (Olssen 2011). We saw this also with five groups of expert participants. The answer and remarks started to be saturated after three participant groups. Selecting a small number of testers must, however, be done carefully because some are better at analyzing and verbalizing their thoughts. Too few participants may lead to unwarranted opinions that could lead design and development teams astray.

Limitations of the research

The primary limitation of this study lies with the number of participants. The test group of junior professionals was too small to get any quantitative data from them; the covid-19 epidemic limited the participant availability for all groups, but

particularly the student game developer group as many were conducting their studies from home.

Also, we decided not to involve anyone who would be inexperienced with either the content or virtual reality technology. In hindsight, they could have represented a fifth group, namely the regular users, in this study. They probably would have had similar challenges with VR technology as the professionals had, as well as the shortage of domain-specific understanding as game developers had. However, this participant group could have brought some new depth to the research.

The survey form should have been tested earlier and refined for a higher validity with regard to the research questions.

All data was first collected and only then analyzed, which is not optimal for constant comparison. The research setup and the questions could have been changed according to the requirements and would have complimented the research better in a way that deeper knowledge could have been extracted during the focus group discussions. In other words, the collected data should have been analyzed before further data generation, and that would have enabled collecting required data that followed clues, filled gaps, and tested interpretations as the study progressed (Chun et al. 2019).

8.1 Recommendations

To succeed in creating virtual and interactive learning environments, based on this research, I recommend the following methods presented in Figure 13:

Clear learning goals, framing, and flow must be created with content specialists with a higher level of domain-specific expertise than the target group. The game development team must validate the concept and prototype with them. Learning goals can be described with visual methods, like storyboards, which are easy to understand for all process stakeholders. Attention should be given to the correctness of high-level assumptions—real world habits and over-enthusiasm to use specific technology must not override the selection of learning experiences to support the learning benefit.

The target user group is suitable for evaluating the user experience, the efficiency of the learning methods, and domain-specific usability. Involving target users supports in-house testing.

The learning outcomes can be tested using several practice scenarios, and the test has to take place in a non-development scenario or environment. Also, testing traditional training as a baseline and comparing the virtual learning environments results brings a realistic picture of the game system's usefulness and readiness.

For usability and user experience testing, target users can be used when the following aspects have been concretized: the learning goals, framing, and interaction. Figure 13 presents the stakeholder resources to their proper use, focusing on how the content specialists should be involved in the process. The process qualifies the optimal timing to use these experts and distinguishes when other stakeholders should give input.

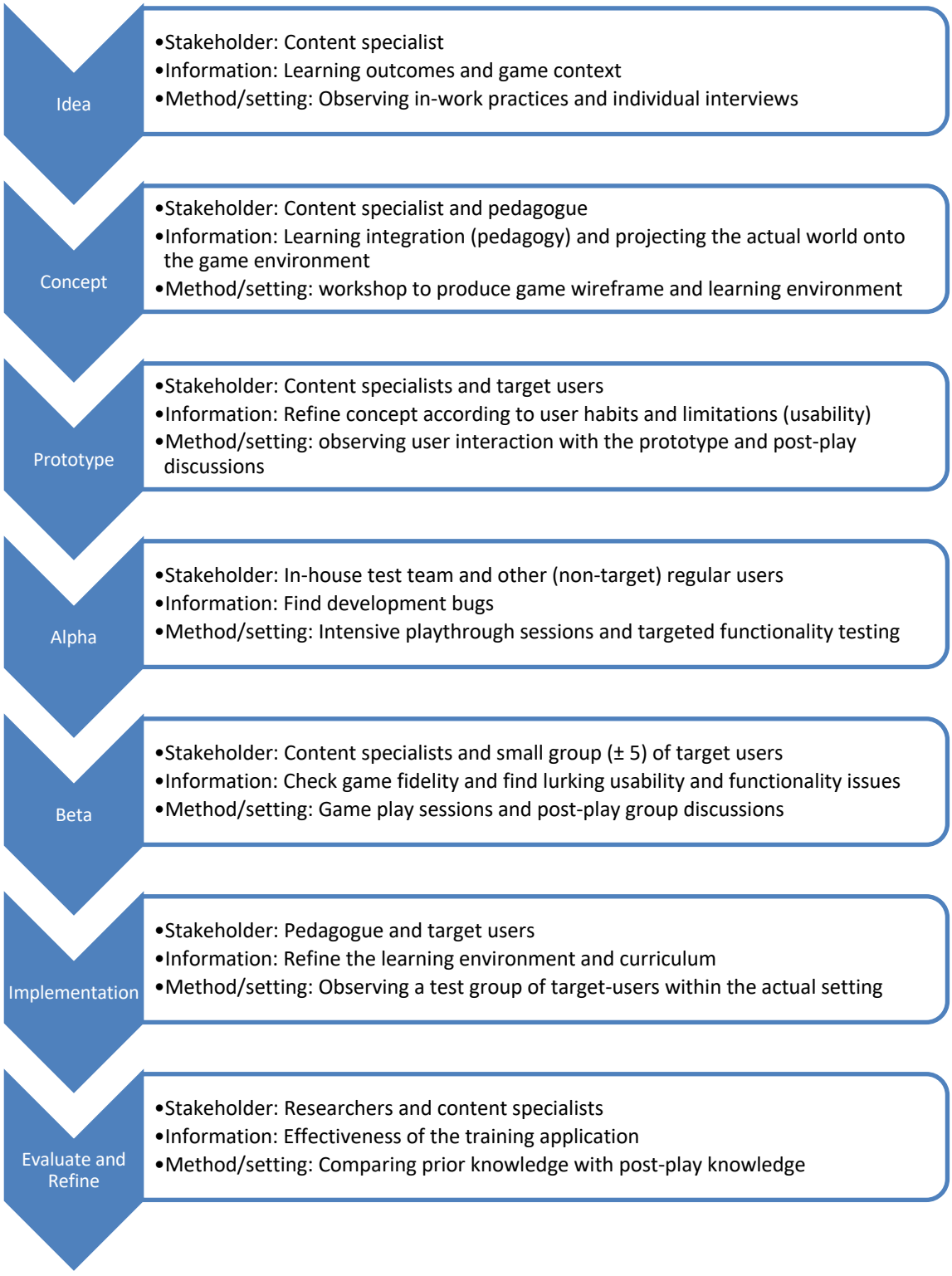


Figure 13: Content specialist involvement in the game development cycle.

9 REFERENCES

Al-Adawi M. Luimula M (2019) Virtual Reality in Fire Safety – Electric Cabin Fire Simulation.10th IEEE International Conference on Cognitive Infocommunications.

Alaraj A. Lemole M. G. Finkl J. H. Yudkowsky R. et all (2011). Virtual reality training in neurosurgery: review of current status and future applications. Surgical neurology international, 2.

Allal-Cherif O. Bidan M. Makhlou M (2016) Using serious games to manage knowledge and competencies: The seven-step development process. Inf Syst Front 18, 1153-1163.

Alexander A. Brunyé T. Sidman J. Weil S. (2005) From Gaming to Training: A Review of Studies on Fidelity, Immersion, Presence, and Buy-in and Their Effects on Transfer in PC-Based Simulations and Games. Aptima, Inc. Woburn, MA.

Alvarez J., Djaouti D. (2010) Introduction to serious game. Paris: Ludoscience. In:

Allal-Chérif O. et al. (2016) Using serious games to manage knowledge and competencies: The seven-step development process.

Anderson L. W. Krathwohl D. R. Airasian P. W. Cruikshank K. A. Mayer R. E. Pintrich P. R. et al. (2001) A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. New York: Longman.

Appelman R. A. (2005) Designing experiential modes: A key focus for immersive learning environments. TechTrends, 49(3).

Ávila-Pesántez D. Rivera L. Alban M. (2017) Approaches for Serious Game Design: A Systematic Literature Review.

Bailenson J. N. Yee N. Blascovich J. Beall A. C. Lundblad N. Jin M. (2008) The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. The Journal of the Learning Sciences, 17(1), 102-141.

Biggs J. B. (2003)Teaching for quality learning at university. Buckingham: The Open University Press.

Billett S. Harteis C. Gruber H. (Eds.) (2014) International Handbook of Research in Professional and Practice-based Learning, Chapter V. Netherlands: Springer Science.

Boije H (2002) A Purposeful Approach to the Constant Comparative Method in the Analysis of Qualitative Interviews. Kluwer Academic Publishers. Printed in the Netherlands.

Bowman DA. McMahan RP. (2007) Virtual reality: How much immersion is enough? in Wohlgenannt I., Simons A., Stieglitz S. (2020) Virtual Reality. Bus inf Syst Eng 62 (5):455-461

Braad E., Žavcer G., Sandovar A. (2016) Processes and Models for Serious Game Design and Development. In: Dörner R., Göbel S., Kickmeier-Rust M., Masuch M., Zweig K. (eds) Entertainment Computing and Serious Games. Lecture Notes in Computer Science, vol 9970. Springer, Cham. https://doi.org/10.1007/978-3-319-46152-6_5

Chun Tie Y. Birks M. Francis K. (2019) Grounded theory research: A design framework for novice researchers. SAGE Open Med. 2019 Jan 2;7:2050312118822927. doi: 10.1177/2050312118822927. PMID: 30637106; PMCID: PMC6318722.

Cowan B. Kapralo B. (2017) An Overview of Serious Game Engines and Frameworks. Faculty of Business and Information Technology, University of Ontario Institute of Technology, Oshawa, ON, Canada

Dalgarno B. Lee M. (2010) What are the learning affordances of 3-D virtual environments? British Journal of Educational Technology, 41, 10–32.

Dowidat L. König J. Wolf M. (2017) The Motivational Competence Developing Game Framework. Lab. for IT Organization & Management, FH Aachen - University of Applied Sciences.

Fowler C. (2015) Virtual reality and learning: Where is the pedagogy? British Journal of Educational Technology Vol 46 No 2 2015 412–422.

Freina L. Ott M. (2015) A literature review on immersive virtual reality in education: state of the art and perspectives. eLearning & Software for Education.

Fullerton T. (2014) Game Design Workshop - A Playcentric Approach to Creating Innovative Games, Third edit. CRC Press.

Gran view inc. Video Game Market Size, Share & Trends Analysis Report, 2020 – 2027. Published Date: May, 2020

Haptic feedback gloves: Haptx.com, Senseglove.com, Manus-vr.com, vrgluv.com 23.5.2021

Ibrahim A. Mahfuri M. Abdallah N. et all. (2020) Using playability heuristics to evaluate player experience in educational video games. Journal of Theoretical and Applied Information Technology. 15th December 2020. Vol.98. No 23

Iuppa N. Borst T. (2010) End-to-end game development: creating independent, serious games and simulations from start to finish. Focal Press, Oxford, UK

Kelly H, Howell K. Glinert E. L. Holding, et al. (2007) How to build serious game. Commun. ACM, vol. 50, p. 44.

Kirkley S.E. Tomblin S. Kirkley J. (2005) Instructional Design Authoring Support for the Development of Serious Games and Mixed Reality Training - Information in Place, Inc.

Lauronen J. Ravyse W.S. Salokorpi M. Luimula M. (2020) Validation of Virtual Command Bridge Training Environment Comparing the VR-Training with Ship Bridge Simulation, July 2020, DOI: 10.1007/978-3-030-50943-9_56, In book: Advances in Human Aspects of Transportation.

Lauronen J. Sakari L. Lehtonen T. (2021) How Simulation Training Can Benefit from Virtual Re-ality Extensions? Case: A Virtual Reality Extension to a Simulated Ship Bridge for Emergency Steering Training. The 12th International Conference on Applied Human Factors and Ergonomics. Conference proceedings. Springer 2021 (to be Published)

Leech N.L. Onwuegbuzie A.J. (2009) A typology of mixed methods research designs. Qual Quant 43, 265–27.

Luimula M. Ranta J. Al-Adawi M. (2020) Demo Paper: Hand Tracking in Fire Safety – Electric Cabin Fire Simulation, In: Proceedings of the 11th IEEE International Conference on Cognitive Infocommunications CogInfoCom 2020, pp. 221-222.

Macqueen K. McLellan-Lemal E. Kay K. Milstein B. (1998) Codebook Development for Team-Based Qualitative Analysis. DOI - 10.1177/1525822X980100020301.

Malone T. W. Lepper M. R. (1987) Making learning fun: A taxonomy of intrinsic motivations for learning. In E. Snow & M. J. Farr (Eds.), Aptitude, learning and instruction (Vol. Volume 3: Cognitive and affective process analysis, pp. 223-253). Hillsdale, NJ: Lawrence Erlbaum.

Marchiori E. J. Serrano A. Del Blanco À et all (2012) Integrating domain experts in educational game authoring: A case study. In Proceedings 2012 4th IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, DIGITEL 2012, pp. 72–76

Markopoulos E. Lauronen J. Luimula M. Lehto P. LaukkanenS. (2019) Maritime Safety Education with VR Technology (MarSEVR). 2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), 2019, pp. 283-288.

Mayes, J. T. Fowler, C. J. H. (1999) Learning technology and usability: a framework for understanding courseware. Interacting with Computers, 11, 485–497.

Mikropoulos T. A. Natsis, A. (2011). Educational virtual environments: a ten-year review of empirical research (1999–2009). Computers & Education, 56, 769–780.

Mitchell A. Savill-Smith (2004) The use of computer and video games for learning: a review of the literature. London, UK. ISBN 1-85338-904-8

Mitgutsch K. Alvarado N. (2012) Purposeful by design?: a serious game design assessment framework. In Proceedings of the International Conference on the Foundations of Digital Games (FDG '12). ACM, New York, NY, USA, 121-128.

Nacke L.E. Drachen A. Kuikkaniemi K. et al. (2009) Playability and Player Experience Research. Proceedings of the IEEE BT - Breaking New Ground: Innovation in Game.

Naranjo J.E. Sanchez D.G. Robalino-Lopez A. et all. (2020) A Scoping Review on Virtual Reality-Based Industrial Training. Appl. Sci. 2020, 10, 8224. https://doi.org/10.3390/app10228224

Oliva D. Somerkoski B. Tarkkanen K. Lehto A. Luimula, M. (2019) Virtual reality as a communication tool for fire safety – Experiences from the VirPa project. In: Proceeding of the 3rd GamiFIN conference, pp. 241-252.

Olsen T. Procci K. Bowers C. (2011) Serious Games Usability Testing: How to Ensure Proper Usability, Playability, and Effectiveness. In: Marcus A. (eds) Design, User Experience, and Usability. Theory, Methods, Tools and Practice. DUXU 2011. Lecture Notes in Computer Science, vol 6770. Springer, Berlin, Heidelber.

Ravyse W.S. Blignaut A. S. Leendertz V. Woolner A. (2017) Success factors for serious games to enhance learning: a systematic review. Springer-Verlag London.

Ravyse W.S. Blignaut A.S. Botha-Ravyse C.R. (2020) Codebook Co-Development to Understand Fidelity and Initiate Artificial Intelligence in Serious Games.

Rutter J. Bryce J (2006) Understanding Digital Games: Sage Publications.

Ryanand W. Charsky D. (2013) Integrating Serious Content into Serious Games, in Foundations of Digital Games.

Saunders J. Bayerl P. Davey S. Lohrmann P. (2019) Validating Virtual Reality as an Effective Training Medium in the Security Domain. Sheffied Hallam University. 2019 IEEE Conference on Virtual Reality and 3D User Interfaces.

Smith, Alexander (2014). The Priesthood At Play: Computer Games in the 1950s. They Create Worlds. Archived from the original on 2015-12-22.

Somerkoski B. Oliva D. Tarkkane K. Luimula M. (2020). Digital Learning Environments - Constructing Augmented and Virtual Reality in Fire Safety, In: Proceedings of the 11th International Conference on E-Education, E-Business, E-Management, and E-Learning (IC4E 2020).

Spanos A (2021) Games of History. Routledge. ISBN 978036735890

Stavroulia K-E Christofi M. Zarraonandia T. Michael-Grigoriou D. Lanitis A. (2018) Virtual reality environments for training and learning authors. In Jensen, L., Konradsen, F. (2018) A review of the use of virtual reality head-mounted displays in education and training. Educ Inf Technol 23, 1515–1529.

Susi T. Johannesson M. Backlund P. (2007) Serious games—an overview. School of Humanities and Informatics, University of Skovde, Sweden. 56. The World Wide Web Consortium (w3C) HTML standard. http://www.w3.org/standards/webdesign/htmlcss.

Tarkkanen K. Lehto A. Oliva D. Somerkoski B. Haavisto T. Luimula M. (2020) Research Study Design for Teaching and Testing Fire Safety Skills with AR and VR Games, In: Proceedings of the 11th IEEE International Conference on Cognitive Infocommunications CogInfoCom 2020 pp. 167-172.

Tong . Sainsbury P. Craig J. (2007) Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007 Dec;19(6):349-57. doi: 10.1093/intqhc/mzm042.

Wohlgenannt I. Simons A. Stieglitz S. (2020) Virtual Reality. Bus Inf Syst Eng 62, 455–461 (2020). https://doi.org/10.1007/s12599-020-00658-9).

Appendix 1: User satisfaction questionnaire

VR fire extinguisher UX survey

* USER ID _____

1. User experinece

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
The application taught me something new about fire safety or acting in a fire situation					
The application helped me to understand how to act in a fire situation easily					
The application helped me understanding how important fire safety issues are					
I prefer learning using VR rather than other leaning methods					

2. User Demographic

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
I enjoyed using the application					
The application was easy use					
The Virtual environment in the application was realistic					
I had a sense of being in the application scenes displayed					

3. User Satisfaction

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
I feel my ability to concentrate/focus has improved					
VR Technology has a positive effect on me					
I would recommend using VR in training Courses in Future					
I feel Safe when Using this Technology					

¶ Section Break (Next Page)

Appendix 2: Informed Consent

INFORMED CONSENT DOCUMENTATION FOR PARTICIPATION IN VR TESTING OF A FIRE EXTINGUISHER

TITLE OF THE RESEARCH STUDY: Do we need to involve expert participants at all levels of testing VR training applications?
PRINCIPAL INVESTIGATOR: Jenny Lauronen
Study supervisor: Mr Werner Ravyse
CONTACT NUMBER: 050 5961698

You are being invited to take part in a research study that is being conducted by Turku University of Applied Sciences research group of . Please take some time to read the information presented here, which will explain the details of this study. It is very important that you are fully satisfied, that you clearly understand what this research is about, and how you might be involved. Your participation is **entirely voluntary,** and you are free to say no to participate. This study will be conducted according to the ethical guidelines and principles of the Decleration of Helsinki and other international ethical guidelines applicable to this study.

What is this research study all about?

This research aims to study the response differences (recorded in the VR environment) between trained fire fighters and regular users not trained in firefighting.

During the research, you will be requested to play a short firefighting scene in VR that involves using a general fire extinguisher. The game will record some of your actions in the scene in the form of metrics that we will analyze. The metrics we collect from your actions will be aggregated and analyzed as a group data set.
This study will involve answering a questionnaire and having a focus group discussion.

Why have you been invited to participate?

You have been invited to be part of this research because you have accomplished fire safety training and over 18 years. Your profile will help us to answer our research question.

What will be expected of you?

You will be expected to:

- Complete this informed consent
- Use a tutorial application to familiarise you with VR headsets
- Use the fire safety application
- Complete a user experience questionnaire
- Participate in a focus group discussion
- Be open to a follow-up discussion at the discression of the principal investigator

The game collects the following metrics:

During the focus group discussion you will be asked to offer your opinions about the application and the learning that you experienced when using this application. The researcher will ask questions to kick start and guide the discussion based on responses from the questionnaires. However, this is an open discussion where you the participant can also bring your own questions for discussion and contribute freely to the topic of discussion. The estimated duration of the focus group discussions is 20 to 30 minutes.

Will you gain anything from taking part in this research?

The benefits for you if you take part in this study will be:

- A direct benefit is that you may add to your existing fire safety knowledge.
- An opportunity to experience VR technology.

Are there risks involved in you taking part in this research and what will be done to prevent them?

There are no foreseen risks. Although the researcher promises anonymity, this cannot be guaranteed. The researcher will, however, use only numbers and letters of the alphabet as name codes for each participant to protect your identity.

Another risk is that although the researcher will try to keep the length of time of the interview discussion bearable, due to its open nature, there is a risk of overrunning its scheduled time. In the event of that happening, the research will call for intermittent refreshment breaks.

There is a chance also that you may feel uncomfortable to talk to researcher about your experience. However, the researcher to conduct the discussions will try by all means to put the participants at ease. And you have the option not to participate in this part of the study.

Foreseen risks are minimal. Rather, there are more gains for you in joining this study than there are risks.

How will we protect your confidentiality and who will see your findings?

In order to protect the participant's confidentiality during the application testing, questionnaire and focus group discussion, you will receive a participant code that you will use on all documentation. These numbers will also be used in all correspondence with the participants.

Who will have access to the data?

Data will be coded to ensure that no link can be made to a specific participant. Reporting of findings will be anonymous by only authorising the principal investigator to have control over the distribution of these findings. Only the study leader and study co-workers will have access to the data and will also sign a confidentiality agreement to protect participants. All data will be password protected. Data will be stored for seven years after which the information will be will be deleted according to the ethical guidelines policy.

What will happen with the findings and results?

The findings and results of this study will be compiled and reported in the Master's thesis of Jenny Lauronen and published in research articles that will be accessible to authorized scholars and future researchers. All data will be assimilated and analysed as an aggregated whole. In this way we ensure that no participant can be identified from their responses.

How will you know about the results of this research?

We will give you the results of this research when we conclude data collection, analysis and reporting by 01/04/2021.

Will you be paid to take part in this study and are there any costs for you?

This study is not funded by any organisation or independent individual outside the research team. The research is purely educational with the researchers not realising any material gains from the research findings. You will, therefore, not be paid for participating in the research. Other than travel (and possible parking fares) to and from the test site, you will not be expected to incur any costs.

Is there anything else that you should know or do?

You can contact Jenny Lauronen, via this number: 050 5961698 or via email: jennylauronen@gmail.com.

You will receive a copy of this information and consent form for your own purposes.

Declaration by participant

By signing below, I ..... agree to take part in the research study titled: Do we need to involve expert participants at all levels of testing VR training applications?

I declare that:

- I have read this information/it was explained to me by a trusted person in a language with which I am fluent and comfortable.
- The research was clearly explained to me
- I have had a chance to ask questions to both the person getting the consent from me, as well as the researcher and all my questions have been answered.
- I understand that taking part in this study is **voluntary** and I have not been pressurised to take part.
- I may choose to leave the study at any time and will not be handled in a negative way if I do so.
- I may be asked to leave the study before it has finished, if the researcher feels it is in the best interest, or if I do not follow the study plan, as agreed to.

Signed at **(*place*)** ..... on **(*date*)** .....

Signature of participant