AALBORG UNIVERSITY
STUDENT REPORT

Electronics and IT
Aalborg University
http://www.aau.dk

Title:
Development of a VR Escape Room for Corporate Compliance Training

Theme:
Master Thesis in Medialogy

Project Period:
Spring Semester 2023

Project Group:
MED10

Participant(s):
Christian Vasileios Tsalidis de Zabala

Lilla Julia Toth

Lilla Julia Toth

Scott James Naylor

Snaylo

Supervisor(s):
Stefania Serafin

Number of pages: 21
Date of Completion: May 25, 2023

Abstract:

This research project focuses on the development of a Virtual Reality (VR) escape room as an innovative approach to corporate compliance training. We investigated theories of active learning, escape rooms for education, game-based learning, motivation, immersion, and flow to answer whether virtual reality escape rooms developed for corporate compliance training are a better alternative to traditional training methods. The study utilizes a quasi-experimental control group with non-probabilistic convenience sampling. It includes a total of n=32 participants, divided equally between the experimental group and the control group. The former experienced the VR escape room training, while the latter received the traditional e-learning self-study training. Data on engagement, motivation, and perceived learning were collected through post-testing questionnaires and direct observation. The metrics used to score the results were the User Engagement Scale (UES), the Situational Motivational Scale (SIMS), and the CAP Perceived Learning scale. The results showed a substantial increase in the affective learning of VR users but no statistically significant difference in cognitive learning. Furthermore, the experimental group surpassed the control group in almost all aspects of motivation with the exception of external regulation, indicating the experience did not feel punitive or forced. The experimental group scored higher in all facets of user engagement, excluding perceived usability. Based on the results, further implementation and polishing of the VR escape room experience are suggested to create a viable product that could offer enhanced engagement, motivation, and learnability for corporate compliance training.

Development of a VR Escape Room for Corporate Compliance Training

Christian Tsalidis, Lilla Julia Toth, Scott James Naylor

May 25, 2023

Contents

1 Introduction

Regardless of the sector or size of an organization, corporate compliance is an essential part of business operations [38]. The education that businesses give their employees about pertinent laws and regulations, business policies and procedures, and their own internal procedures and policies are known as compliance training [45]. It is crucial not only because it is oftentimes required by law to give employees explicit training on internal policies and local regulations [3], but also because it is in the best interest of organizations to have a compliant and capable workforce. Compliance violations can result in penalties, fines, lawsuits, damaged reputation, and damaged finances, and can cause businesses to underperform. Therefore, effective compliance training to avoid employee wrongdoings is key in order to mitigate risks, avoid costs, raise awareness, improve business outcomes, and increase the competitiveness of a company.

Compliance training programs can cover a wide variety of topics, such as Data protection & privacy, Transparency & anti-corruption, or Ethics & integrity [23]. Not all compliance topics apply to all organizations, some are solely relevant to companies in particular sectors. The person responsible for compliance of a company keeps track of what training is mandatory or relevant [31].

Traditional compliance training methods include facilitated training using case studies, web-based, e-learning, PowerPoint lecture approach, or talking head videos [40]. According to a survey published in 2019, almost half of the people (49%) who receive training through traditional methods find it boring [4]. The same survey found that 28% of the people scored low on motivation and 12% found them unproductive. According to another research conducted on self-study methods, such as e-learning and other traditional learning management systems, 15% percent of participants said they just click through without reading or listening, and 34% only skim-read the material or does not fully pay attention to audio and video content [10]. All these statistics point to the fact that these methods are not engaging enough. The more engaging methods, such as facilitated training can be quite costly, while e-learning methods are more cost-efficient as they can be registered ”en masse”. The main issue with disengaging training is that it results in low motivation and poor knowledge retention [12], which in turn increases the chances of compliance breaches [53]. There are methods that can make compliance training more engaging, such as gamification and game-based learning.

Over the past years, Escape Rooms (ERs) emerged as a popular recreational activity worldwide [29]. They can be characterized as live-action team-based games where players face various challenges to carry out a given mission in a limited amount of time. Originally the mission was to get out of a room or set of rooms, hence the name “Escape Room”. However, as of late, these missions might be anything ranging from murder mysteries to rescuing hostages [34]. Because ERs are engaging leisure activities and they also promote beneficial skills such as cognitive competence, teamwork, leadership, critical & creative thinking as well as communication, they have been adapted in educational settings [29]. These educational ERs, most of which are conducted in a traditional physical way, have been applied in primary, secondary, and higher education as well as professional development programs with success in teaching a broad range of disciplines. The issue with traditional physical ERs is that their designs are limited by their material aspect [50]. Beyond this issue, if companies were to use ERs as compliance training tools it could be quite costly to set them up on their own premises or to transport large amounts of employees to established ER locations; as well as time-consuming to reset the room after every use.

The concept of an educational ER is rather new, and the existing research is mostly focusing on the effect and application of physical escape rooms [50]. There seems to be little exploration into the effect of virtual educational escape rooms. Over the past years, Virtual Reality (VR) has been rapidly developing, becoming more attainable, attractive, affordable, and portable [17]. An Educational ER implemented in VR could potentially tackle the issues of limited design options, lack of space, and high costs, making it a viable option for corporate compliance training. Additionally, due to the immersive nature of VR, it could deepen engagement and raise motivation, resulting in better knowledge retention [53] [25] [11].

This paper aims to investigate the current landscape of educational escape rooms, including common practices in their design processes and the underlying theories on which they are built. The resulting findings should serve as a set of requirements in the design of an educational VR prototype, which will then be evaluated and compared against a traditional training method.

2 Analysis

2.1 Compliance Training

In the modern marketplace, all business sectors are heavily governed by laws and regulations. Additionally to abiding by the relevant laws, organizations are widely advised to develop and adopt a code of conduct to prevent unwanted behaviours [26]. This code generally includes rules, responsibilities, principles, values, and guidelines for appropriate and ethical behaviours [39]. It is the responsibility of organizations to make sure that their employees follow the law and the code of conduct to avoid adversity that may harm the integrity of the company. A compliance program is the control system through which employers ensure the proper behaviour of workers. A code of conduct is usually the starting point of an organization's compliance program, and training is the primary way to help people understand and apply this framework [47].

Training can be delivered through various methods. Instructor-led or facilitated training involves the employees attending classes and workshops where they are delivered the training materials. If done correctly it allows for collaboration, discussion, and active participation for employees. It can be an effective learning method but it requires more time investment from employees and financial investment from the company [40]. Oftentimes, these kinds of training turn into passive learning, where participants sit through long presentations and participate in forced discussions. While it is an engaging method in itself, the underlying motivation for attendance is because of external factors.

E-learning training methods administered through learning management systems require less time investment from employees and incur fewer costs for companies [54]. Participants can learn at their own pace with the help of informational videos, text materials, and sometimes interactive elements like quizzes. While more economical, the trade-off is that it is much less engaging than facilitated training, and it is still mostly motivated by external factors [40]. With participants skimming through materials or switching tabs in their browsers during videos and then guessing the answers for the questions this method could also be categorized as a form of passive learning [19] [10]. On top of this, through e-learning self-study training participants are even deprived of the chance to learn collaboratively.

2.2 Escape Rooms

The large success of traditional recreational ERs in the past decade can be attributed to their reliance on immersion [42] [52]. ERs can have a vast variety of themes and scenarios [33]. The goal of these ERs is to take the participants from their everyday life and transport them into the context of the game [50]. They try to achieve a deep state of involvement and engagement by mentally and emotionally engrossing participants through the narrative, environment, sensory stimuli, etc [52] [9] [25]. The narrative must be coherent and consistent with the environment and other sensory stimuli and participants must be able to relate to it to prevent cognitive dissonance [33].

Another vehicle for engaging and immersing participants is the puzzles. All forms of challenges are referred to as puzzles in ERs, and they generally all follow the same structure of a challenge that requires a solution that provides a reward when solved. Rewards can be, for example, clues for the next step or for another puzzle. Most often puzzles in ERs require the cognitive skills of participants. They can also be puzzles that require physical skills, such as dexterity. Puzzles of ERs can be organized in different ways. In a sequential structure, there is a starting point, or puzzle, that unlocks the next puzzle, and so on. With an open structure, there is no starting point, and puzzles are independent of one another, but they all contribute to the solution of the final puzzle of the game. These types of puzzles are called meta puzzles. A path-based structure is essentially multiple sequential puzzles independent of each other besides leading to the same final point. Participants may start at any of the paths of sequential puzzles and end up at the same final puzzle. The last structure is a combination of the previously mentioned structures, resulting in a hybrid [34] [52].

2.3 Escape Rooms for Education

According to the systematic review conducted by Vedkamp et al, ERs have been used in a wide variety of educational purposes. These applications include recruiting students, onboarding to institutional services and emergency situation training. Other papers included in their review employed ERs as a research environment to obtain data about students' information search behavior, learning processes in groups, or how teamwork and leadership skills are utilized between students. The development of ERs by students to improve design skills is another example included in their study. Finally, ERs to advance expertise or knowledge particular to disciplines (e.g.: computer science, medicine, mathematics) or to aid the development of generic competencies are mentioned [50].

Educational ERs employ the same structure as recreational ERs, where teams must complete challenges in a limited time that require both physical and mental engagement. Designers of educational ERs aim to develop relevant and authentic environments with meaningful activities oftentimes by taking and adapting material from established leisure ERs. A key difference between recreational and educational ERs is their target group. Recreational ERs are built to appeal to a broad audience, and they often pride themselves on their low success rate, to further draw people in by posing a difficult challenge. Educational ERs on the other hand are developed for a particular target group with well-defined learning goals based on curriculum [50].

Contrary to recreational ERs' low success rate, educational ERs' are created to attain a high success rate so that all learning goals are met, and students can feel accomplished and have a positive experience [50]. This puts designers of Educational ERs into an especially difficult position. Puzzles and challenges of the game have to be created to be difficult enough to cognitively engage participants but not too difficult to cause frustration and unsuccessful escape attempts. Additionally, the puzzles should align with the curriculum to actually deliver the learning goals [29]. Educational ERs might also not have sufficient space available (i.e.: multiple connected rooms) to implement more complex experiences. The lack of space poses an additional challenge when taking into consideration that educational ERs also try to cater to larger amounts of people at once, as opposed to recreational ERs which most often operate with smaller groups. Both traditional educational and recreational escape rooms are limited in their physical aspect, meaning that the puzzles must utilize available material resources, and the outcomes have to be in a format that is usable in the following challenges of the game [50].

Depending on the learning goals defined for ERs they can be integrated and positioned in the educational environment differently. Learning goals related to content knowledge and related skills are most often to foster, demonstrate, extend, integrate or assess said knowledge and skills. Other learning goals are related to more general skills, such as teamwork, communication, problem-solving or critical thinking skills. Educational ERs can help students develop or practice the aforementioned skills. The last category of learning objectives most often employed by educational ERs are affective goals. This domain involves the emotions, underlying values and attitudes that have an effect on the learning process and behavior of students. An example for this domain could be a learning goal that requires students to demonstrate the capabilities of working under pressure. Such a goal would be a common requirement in many medical fields [50].

Depending on the defined learning goals educational ERs can be positioned differently in the curriculum. They can be used to introduce students to new subjects or content, or in addition to lectures to help students demonstrate stronger content knowledge and related skills. They can be used as a type of formative assessment, or as a standalone activity. At their core educational ERs are basically a type of educational games, and as such their mechanism of action can be explained by theories of Game-Based Learning [50].

2.4 Game-Based Learning

Digital games promote active participation and encourage involvement which makes them inherently engaging activities that can provide players with a sense of enjoyment and immersion [25] [2]. Challenges of games usually demand problem-solving, strategic planning, critical thinking, and other cognitive skills that involve the same higher-order thinking required in the learning process [43] [7]. The outcomes of challenges often provide immediate feedback to players, which can also be a valuable element, as it can speed up the learning process [50]. When applied in an educational setting through these properties they can be more effective in engaging students [13].

Game-based learning (GBL) can be defined as using the characteristics and mechanics of games to build an engaging and immersive learning experience for delivering specified learning outcomes [13]. Essentially, the game becomes the medium for the curriculum or parts of it as opposed to gamification, where certain game elements are applied to non-game activities [22]. GBL can be both non-digital such as ERs, board games, and card games, and digital such as video games. GBL aims to tap into the intrinsic motivation of learners, by giving them control over an immersive environment that appeals to their curiosity and they are challenged. Providing players with immediate feedback when overcoming challenges and progressing through the game provides recognition and a sense of achievement. Beyond challenges, goals, feedback, and rewards, games can also heighten intrinsic motivation through competition or cooperation. Within the engaging and dynamic environment of the game, learners can make decisions and experience the consequences of their actions [20].

The foundation of GBL is based on the active learning approach of constructivism, developed by Piaget et al [7]. Bonwell and Eison [6] define active learning methods as *"instructional activities involving students in doing things and thinking about what they are doing"*. There can be a number of activities that are included in the above definition, but it is important to note that they should require higher-order thinking from the learner. Active learning strategies often involve making connections between new information and current mental models of learners or presenting them with misconceptions so that they may challenge and reconstruct their existing mental models. Games provide an excellent framework where such activities and strategies can be implemented [7].

Working together in the context of a game to achieve is another approach rooted in constructivism that enables active learning. Collaborative learning can be described as a process in which groups of 2-6 people work together to share their abilities, knowledge, and contributions. Vygotsky argued that due to the innate collaborative nature of learning it cannot be separated from its social context [51] [48]. Activities that rely on teamwork can facilitate learning, through learners solving issues beyond their individual capabilities with the help of their peers. This way mental models of students must be made explicit so that they may be subject to discussion and are not only challenged by initial instructional activities but also by their groupmates. Additionally, the members of the group extend each other's mental models as a form of distributed cognition. Through the network of interactions that social discourse entails knowledge emerges, and cognitive functions are developed. What makes collaborative learning even more powerful is the effect that learners have on one another in terms of motivation and self-esteem. If positive interdependence of members is fostered, individuals demonstrate higher self-efficacy and increased motivation.

2.5 Motivation theory

Self-Determination Theory (SDT) is a psychological framework developed by Deci and Ryan that focuses on the interplay between psychological needs, motivation, and well-being of human beings in various situations [14]. The theory proposes that individuals have three basic needs psychological needs, and the satisfaction of these can result in positive outcomes in a variety of contexts. The enhanced motivation and well-being stemming from becoming self-determined by fulfilling these basic needs can for example positively impact learning outcomes.

Autonomy refers to the basic need of an individual to have control over their own actions and have a sense of choice. In an educational setting, autonomy can be achieved if learners are given the ability to take control over their learning or are given choices regarding the structure and method of their education. In a game context, autonomy can be experienced if players have the means to discover, shape and explore the virtual world.

The need to feel capable and effective when dealing with challenges is referred to as **competence**. In an educational environment, competence can be promoted by ensuring that learners are shown their progression in mastering content knowledge and related skills. To aid the improvement of learners educators might supply them with necessary help in their progression of becoming competent in a given area, or provide feedback on their improvement. In a game context, competence is the same need for displaying mastery as in education, but games are equipped with different tools to convey this to players. Often times games use progress bars and other visual cues to keep players informed on their achievements.

Relatedness is the need of an individual to be connected with others and to have a sense of belonging. In an educational context, relatedness comes from establishing social connections with peers and working collaboratively on achieving learning outcomes. In a game context, playing in cooperation or competitively in the same virtual environment is the primary source of relatedness.

When all of the above-mentioned components of SDT are fostered successfully, individuals are more likely to experience **intrinsic motivation**. This type of motivation refers to the internal desire of people to engage in activity. Increased intrinsic motivation enhances engagement which makes it a desirable goal to achieve in both an educational and gaming context. Additionally, within education, it is also linked with improved learning outcomes. On the other end of the spectrum, there is **extrinsic motivation**, which comes from engaging in an activity because of external factors, such as rewards or punishment. Excessive external pressuring factors can result in poor performance and a negative mental state, making it an undesirable effect [24].

2.6 VR

Virtual Reality (VR) has been rapidly developing in the past decade, becoming widely available to the general public and it is now used in a multitude of ways, such as immersive gaming, medical applications, and training simulations for high-risk jobs [18] [49] [17]. Because of the many ways it can be utilized, it has also been a popular research topic in the field of Human-Computer Interactions.

VR has the ability to draw users in by allowing them to interact with highly immersive computer-simulated environments. Brown and Cairns found that immersion in this context can be understood as the degree of involvement with the VR-generated environment [8]. They identified three levels of immersion starting from the lowest level these are engagement, then engrossment, and finally total immersion. During the first level, players attempt to overcome initial barriers such as learning the controls and understanding basic game mechanics. If these elements seamlessly become natural to the individual, and the environment also emotionally affects them, they can become engrossed. If they become even further involved and experience a sense of presence by forgetting about the outside world they achieve total immersion. This state requires intense focus and attention from players.

This definition of immersion overlaps with the state of flow proposed by Csikszentmihalyi [32]. The experience of flow occurs when participants are subjected to the process of optimal experience, and become so caught up in it that it becomes their sole area of focus. This theory presents 8 components of flow, that serve as important design considerations for a successful VR experience.

The two above-discussed theories suggest that achieving these states can stimulate the intrinsic motivation of individuals by presenting them with an activity they can participate in for the inherent satisfaction it provides. Based on this, VR educational ER experiences designed around compliance training goals to facilitate immersion could potentially offer an engaging alternative that intrinsically motivates people to participate and achieve better learning outcomes in an active and collaborative way. Contrary to traditional physical ERs, a VR version is not limited by physical and material aspects. Due to the inherent immersive nature of the technology, it allows for more possibilities in the design of experiences, while avoiding costs associated with materials and the time investment required to reset a room.

2.7 Hypotheses

This research aims to investigate whether the Virtual reality compliance training experience developed for this project has any effect on the engagement, motivation, and perceived learning of participants compared to popular traditionally used training methods, such as e-learning self-study training. This statement can be broken down into three parts corresponding to engagement, motivation, and perceived learning respectively. They are formulated as follows.

Engagement:

> $\mathbf{H0_a}$: Virtual reality escape rooms for compliance training have no effect on the engagement of employees compared to the traditional e-learning self-study method.
>
> $\mathbf{H1_a}$: Virtual reality escape rooms for compliance training positively impact the engagement of employees compared to the traditional e-learning self-study method.

Motivation:

> $\mathbf{H0_b}$: Virtual reality escape rooms for compliance training have no effect on the motivation of employees compared to the traditional e-learning self-study method.
>
> $\mathbf{H1_b}$: Virtual reality escape rooms for compliance training positively impact the motivation of employees compared to the traditional e-learning self-study method.

Perceived learning:

> $\mathbf{H0_c}$: Virtual reality escape rooms for compliance training have no effect on the perceived learning of employees compared to the traditional e-learning self-study method.
>
> $\mathbf{H1_c}$: Virtual reality escape rooms for compliance training positively impact the perceived learning of employees compared to the traditional e-learning self-study method.

Joining the above three statements the null and alternative hypothesis of this research are:

> ***H0****: Virtual reality escape rooms for compliance training have no effect on the engagement, motivation, and perceived learning of employees compared to the traditional e-learning self-study method.*
>
> ***H1****: Virtual reality escape rooms for compliance training positively impact the engagement, motivation, and perceived learning of employees compared to the traditional e-learning self-study method.*

3 Concept & Design

3.1 Context

Corporate compliance training can encompass a large number of topics. To determine an appropriate topic and synthesize a list of learning goals, existing available policies and training curriculums of Danish corporations were investigated [1] [28] [37] [46]. Ethics policies were found to be one of the most common parts of compliance programs.

The main purpose of including a code of ethics in compliance training is to provide employees with a set of rules and expectations that they can follow to avoid engaging in unethical behavior, violating laws, causing harm, or engaging in wrongdoings in any other way [26]. The identified existing policies were synthesized into learning goals with the help of Bloom's Taxonomy [5]. The main topics within ethics compliance training are anti-bribery and corruption, gifts and hospitality, conflicts of interest, data protection and confidentiality, human rights and respect, competition law compliance, political activities, sponsorship and donations, external communications, and handling ethical social dilemmas. Each of these areas is made up of the policies of the company regarding the topic, as well as guidelines for employees on how to conduct themselves as representatives of the company in situations relating to the aforementioned areas.

The aim of the escape room was to fulfill the learning goals found in appendix B. The main motivation was to increase the engagement, intrinsic motivation, and perceived learning outcomes of participants so that they can be better equipped to deal with issues arising from the previously listed topic. The escape room was positioned to be a stand-alone activity, with the objective to foster content knowledge and related skills, as well as developing teamworking skills. The participants were expected to acquire new knowledge, without any prior preparation, that involved the use and development of their inner moral compass [30]. The escape room had a briefing session, that provided an introduction to the story, the controls of the game, main objectives, and an initial clue. There was no debriefing session after the escape room.

The target group for whom this VR educational ER was developed can be described as any individual who is employed at a business where corporate compliance training is relevant. Sectors where corporate compliance training is necessary can be for example Finance or IT. This encompasses a large number of different people, and it is difficult to narrow down to a specific set of people who share more common attributes. Based on data from Statistics Denmark [15], the largest segment of people working at public and private corporations are between the ages of 19 to 65. This range includes multiple generations with different characteristics. Younger generations are more comfortable with technology, while older generations might not be as adapted. Designing for such a large target audience means that the various elements of the game, such as controls, mechanics, challenges, etc. have to be clear and understandable for individuals from multiple generations.

3.2 Design

The design of the VR educational ER was based on the concepts, theories, and approaches investigated in the analysis. Additionally, to the findings of the previous chapter, the design also utilized the VR PLAY guidelines [16], which is a set of 33 principles organized into 5 categories of heuristics with the purpose of aiding the development of more usable and playable VR experiences.

The VR educational ER was developed as an experience with an immersive narrative [Section 3.3] and an environment where participants had to solve a number of cognitive and physical puzzles in a limited amount of time. The experience was designed for 4 participants at a time, located in the same physical and virtual space where players were free to move around. Locomotion in the shared space was done by physically moving around, instead of methods like teleportation or smooth locomotion using the VR controllers. This type of motion was also a way to reduce complexity in order to appeal to the target audience since no additional controls were necessary. Players simply had to continue moving around as they would in the physical world. Because the experience takes place in VR, except for the initial brief on the narrative, the only material requirements were the VR headsets and a 4 meters by 4 meters empty space.

Throughout the design process, the project team had two sessions with the CEO of Escape Copenhagen [1] to gather information on a number of topics regarding the design process of an escape room. The sessions provided the team with insights into difficulties, key areas of attention, them-

[1] Escape Copenhagen `https://escape-cph.dk/en/`

ing, timing, and organization of recreational escape rooms. During the second session, the project team also tested one of Escape Copenhagen's unreleased rooms, which was specifically developed for corporate team-building events, and included educational elements regarding cybersecurity. This field study allowed the project team to gather information from the perspective of the users on the different types of puzzles and their structures, and how a professional escape room experience is conducted from start to finish.

The puzzles in the room were organized in a sequential structure. This structure was selected for multiple reasons. First, to encourage participants to engage in the same puzzles at the same time, and thus all fulfilling the learning goals. Ensuring simultaneous participation also supported the fulfillment of the need for relatedness as suggested by the SDT. Second, to ensure that the learning goals are structured in a logical consecutive order. Increasing complexity and difficulty in a sequential structure is also easier from a development point of view to support the experience of flow. Third, a sequential structure leaves less room for error and requires less guidance or intervention resulting in a higher success rate, and a fulfillment of all learning goals covered in the experience. This structure is also easier to track for the person administering it, allowing for better guidance. It also allows for a more apparent perception of progression for all of the participants to support the need for competence. The cognitive puzzles of the game are the main mechanisms through which the learning goals are met, while the physical puzzles are additional ways to enhance immersion and engagement.

Preventing frustration caused by getting stuck at any point in the game is important to keep students engaged. Getting stuck can also result in a lower success rate, which means that some learning goals remain unfulfilled. Both Educational and Recreational ERs tackle this issue by providing hints [50]. Participants were provided hints verbally by the conductor of the escape room if they asked for it, or if the conductor deemed it necessary. There were no limits on hints in this experience.

To encourage collaboration, the experience integrated unique player skills. These skills allowed players to interact with or see elements of the game that other players could not. As the experience was made for 4 participants, there were 4 separate unique skills. Participants were informed during the initial brief that they will have certain abilities, and their specific roles were revealed once they began the experience. The skills were added to aid in fulfilling the need for autonomy, competence, and relatedness. They served as a starting point for collaboration while ensuring that each participant gets to have an effect on the outcome of the game to achieve a sense of agency and the feeling that they are capable members of the team. The unique player skills were constructed to fit into the narrative of the game and to be relatable for all members of the target audience. The skills and their descriptions can be seen in table 1.

Player Role	Skill
Electrician	Modify the flow of electricity
Programmer	Interact with computer interfaces more easily
Cryptographer	Decrypt encrypted codes
Data analyst	View correlations between data

Table 1: This table describes the types of roles each player was assigned, along with their skill.

Traditional recreational escape rooms usually last one hour [50]. A typical e-learning compliance training lesson on average ranges between 5-30 minutes [55]. To make the training a desirable product for companies while still giving employees an engaging experience, it was decided that the VR educational ER experience should not be longer than 30 minutes, excluding the initial brief. This also aligns with the recommended amount of time to spend in VR by Meta, the manufacturer of the headset used in this project to avoid fatigue and other possible side effects that may arise from prolonged usage of VR [36]. Implementing and evaluating a complex multi-room educational ER experience is a time-consuming task, therefore it was decided that only the first room would be developed and tested. The aim was to make this section of the game last about 15 minutes and include at least 2 learning goals. After establishing the target group, learning goals, duration, puzzle structure and types the narrative of the ER was developed as well as the details of each puzzle.

3.3 Narrative.

The narrative is an important design aspect, as it serves as a vehicle to provide meaningful context. The context within the game should be relatable to the real-world application of the knowledge or skill delivered. This way the learning process is facilitated by allowing participants to connect the knowledge delivered by the game to practical situations. If the narrative of the game is well-built, and coherent with the environment, characters, challenges, and interactions the players will be immersed and thus more engaged [50] [25]. Besides aligning game mechanics with the narrative, educational ERs also have to line them up with learning goals [29]. Mechanics should provide learners with opportunities to obtain, practice and apply learning goals while remaining consistent with the narrative [50]. The challenges that deliver the learning goals within the narrative should provide an appropriate level of difficulty. They should be engaging and stimulating for participants without being too easy or too difficult. Players should have clear goals, and progress through the game with the puzzles providing clear and immediate feedback to help induce a state of flow that can positively impact learning outcomes [32].

3.3.0.1 Theme.

The theme of the VR educational ER of this paper was chosen to be a combination of two popular concepts: Sci-fi and Mystery. The Sci-fi trope allowed the seamless incorporation of the VR headset, while the mystery theme was used to construct a storyline that prompted users to investigate in-game events and hunt down shocking truths.

3.3.0.2 The story.

The introduction to the story given to the participants before they put the VR headset on can be seen in Appendix C. In this introduction, the players were called upon as the *"company's"* most valuable employees, each an expert in the field of ethics. They were informed that the Artificial Intelligence developed by their company for detecting ethics violations has found some serious wrongdoings within the company. Before the AI could provide more information, it was hacked, and the entire system of the company was inaccessible through traditional methods. They only had 15 minutes before all the data in the system was destroyed by some mysterious hacker. As an alternative access method to the company's network, they had their consciousnesses uploaded to the cyber world where they had to retrieve two major ethics violations, scan these, categorize them, and relay them to the outside world. Once they put the VR headsets on their consciousnesses would take the form of a virtual avatar in the cyber world, where they found themselves in the e-mail directory of the company with certain unique abilities. The details of the unique abilities were not revealed, and the particular roles were relayed to participants as the first thing when they put the headset on. This was done to encourage players to immediately start communicating and collaborating in figuring out what their roles were and what abilities they might provide. Beyond enhancing collaboration and communication, the abilities also served as a way to provide additional competence and control in the virtual environment. The initial clues given to the participants were that there are two major ethics violations and that the Artificial Intelligence traced these starting from an anonymous e-mail sent to a person named Hannah Johnson at Human Resources in November 2053. Participants were also told instructed there were two different types of controls used in the game, one for grabbing objects and one for pointing and selecting. To reduce the complexity from which issues may arise the controls were kept simple so that the larger target group from various age groups may be able to participate in the experience with fewer barriers. This introduction aimed to give the players a clear initial goal, and the means of control over their environment to achieve autonomy. Most interactions and manipulation of the environment are based on some form of real-life experience. For example, physical puzzles require twisting of knobs, opening cabinets, or pressing buttons. Besides a clear goal, a sense of urgency was established by the active hacker in the system who was in the process of deleting all the data. If a full experience with multiple rooms was implemented, additionally to the initial goal of finding the ethics violations, the players would also have to try and catch the hacker and the total time would be extended to 30 minutes.

3.3.0.3 Puzzles.

Once in the starting room, which represented the e-mail directory section of the high-tech internal world of a computer network the players were free to interact with their environment. The room had odd machinery and a bustling digital metropolis outside the window as shown in figure 1. The city outside the room represented other parts of the cyber world, with references to well-known websites and applications.

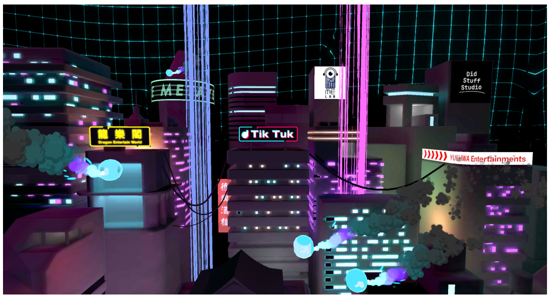


Figure 1: Figure illustrating the cyber world city from the perspective of a player situated inside the room looking outside the window.

In the room, there are multiple interactable elements, listed in table 2 with their respective functions. Table 1 shows the type of player roles along with their unique abilities. Above the window, a display indicated that there were 0 out of 2 violations identified, as seen in figure 2. This element was there to reinforce and clarify the goal of the game and to decrease the cognitive load. Based on their initial clue, players had to try to power on the mail fetcher to look for the anonymous e-mail mentioned in the introduction. To do this, the electrician had to open the cabinet and solve the physical pipes puzzle. If any other player tried to interact with the cabinet or its contents their hands would be pushed away, and sparks of electricity appeared signifying that this section required a player with a different skill. In this puzzle, the electrician had to redirect the flow of electricity from the entry plug to one of the plugs of the machines in the room. The plugs were labeled with post-it notes, with the names of the machines on them. This puzzle can be seen in figure 3. While only the electrician could interact with this puzzle, the other players could aid them by giving suggestions on what elements to rotate to get the flow of electricity to the correct plug. The unique skill of the electrician should meaningfully aid in shaping the outcome of the game and provide the players with competence and autonomy.

Figure 2: Figure illustrating how players were reinforced with the goal of the game. This text also served as a feedback channel to the users when a certain object was unlocked, as well as for violation reports conditions.

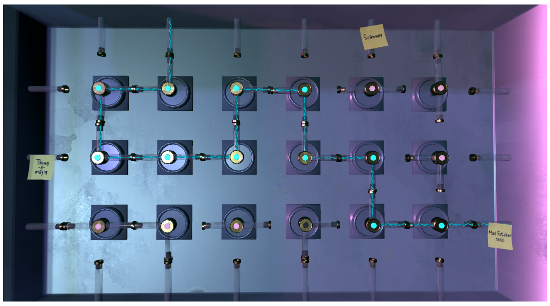


Figure 3: Figure illustrating the flow of electricity in the circuit box, with which only the player with the electrician role could interact with to power on the appropriate objects.

If the flow of electricity is successfully redirected to the mail fetcher, it will power on, which is indicated by both a sound and the screen of the machine turning on. Both of these sensory stimuli serve as a form of feedback to notify players on their success and progression in the game. Figure 4 shows players trying to figure out how to interact with the mail fetcher. Players can press the button on the machine to initiate a dump of all the emails from

Element	Function
Thing-e-majig	Enables the creation or destruction of objects that are essential for player interaction within the game.
Circuit Box	Provides electrical power to various machinery and devices that players need to interact with in order to progress in the game.
E-mail Fetcher	Allows players to retrieve specific emails from designated dates on the company's intranet, providing them with important information for solving puzzles and advancing the narrative.
Scanner	Enables players to scan various objects in the game environment, helping them gather information and identify any potential violations or anomalies.
Flagging System	Provides players with a mechanism to report any violations or suspicious activities they encounter during gameplay, fostering a sense of accountability and promoting ethical behavior within the game world.

Table 2: This table shows which types of elements users could interact with, along with their corresponding functions.

November 2053. They then have to search through the e-mails, to find the one mentioned in the brief for the game. E-mails can only be read by the cryptographer, for everyone else, the emails are just a random mix of letters, as shown in figure 5.

Figure 4: Figure illustrating two players wondering how to interact with the mail fetcher in front of them.

The sender and recipient of the e-mails can only be seen by the data analyst. This is shown by connecting the e-mails with a line of data to the wall of employees in the room, as can be seen in figure 6. All e-mails are connected to two points on the wall, except the anonymous e-mail. Once again, the unique skills encourage collaboration and satisfy the players' need with that given skill for competence and autonomy.

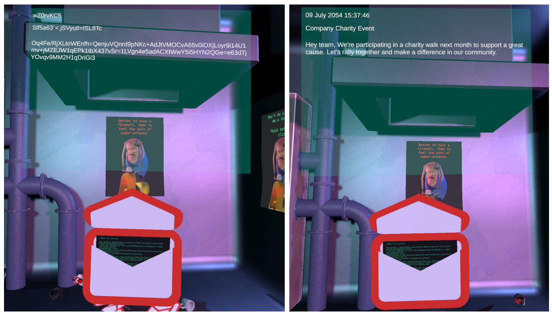


Figure 5: Figure illustrating how players read the e-mails. All the non-cryptographer players read gibberish text (left), whilst the cryptographer can read and understand the actual contents of the e-mail (right).

Figure 6: Figure illustrating how the data analyst player can see lines indicating both the sender and recipient of the e-mail.

The anonymous e-mail contains an ethics violation as well as a clue for the next puzzle. In this e-mail, participants find out about discrimination happening in the hiring process of the company, and some clues to a potential act of bribery involving Jasper Kim and a person named Zachary. To flag the email as an act of ethics violation, the electrician must power on the scanner, where players must place the e-mail. After scanning, they have to categorize the violation. A screen appears on the window of the room, with a list of possible ethics compliance topics, as shown in figure 7. Through this UI which is referred to as the "flagging system", the players must vote on one of the topics which they think fits the nature of the violation best. All players must agree on the topic chosen. If voted incorrectly the option will flash red followed by the sound of a buzzer indicating a wrong answer. If voted correctly, the option will flash green, and a sound indicates a correct answer. The screen then displays a list of possible guidelines regarding the violation with an indicator in the top right corner displaying how many options have to be selected to progress. Once again, all players have to vote, and they all have to agree on which guidelines are applicable. In the corner of each tile, there is a counter that displays how many people have voted for that option. The tiles are not placed in the same place for all players, to further encourage discussions amongst team members. This puzzle corresponds to the learning goal regarding human rights. If all the correct answers are selected, the screen will disappear and the display above the window changes to 1 out of 2 violations detected signifying successful progression in the game. This display will now also say mail fetcher wheel unlocked.

The anonymous e-mail contains clues to the next puzzle. Participants can gather from this e-mail, that the events referring to the potential bribery happened a month before November 2053. To adjust the mail fetcher for accessing e-mails from October 2053, players must create the now unlocked missing wheel with the help of the Thing-e-majig. After rerouting the power to this machine, the programmer can create the wheel through a dot-connecting mini-game. In this mini-game, the lines

Figure 7: Figure illustrating the flagging system. By selecting the individual panels, players could vote for which policies were appropriate to the currently detected violation.

that connect dots are chased by bugs, and if a bug reaches the line it is severed. In theory, every participant can operate this machine, but the number of bugs is much higher than they are for the programmer, and they are also much faster.

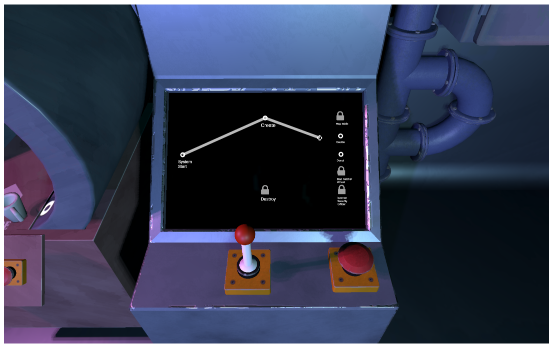


Figure 8: Figure illustrating the computer (Thing-e-majig) powered on to be used by the programmer.

With the wheel created, and the mail fetcher re-powered, the e-mails from October 2053 can be accessed. Once again, the cryptographer can read the mail, while the data analyst can see the sender and recipient. The correct e-mail will be connected to Jasper Kim and Zachary Kim, as visible to the data analyst, and it will contain further proof that bribery is being committed by these people. The e-mail also includes a map code to a virtual meeting room, where the bribery was committed. The e-mail has to be flagged, by powering on and using the scanner. Once scanned, the same system of ethics violation categorization (flagging system) will show up on the window. Players have to discuss and successfully identify the category and the relevant guidelines. This puzzle corresponds to the second learning goal regarding corruption.

If the violation is successfully reported, the display above the window changes to 2 out of 2 violations identified, and it will also inform players about the map table being unlocked. The map table can be created by the programmer through the powered-on Thing-e-majig, as seen in figure 9. In the map table, the data analyst has to connect data points to get the team transported to the virtual meeting room. The data analyst can see different colored data points on the map table, while the rest of the team can only see grey points. The cryptographer has to translate the earlier obtained virtual conference room map code into a series of colors, with the help of the decryption guide found rotating at the bottom of the map table. For everyone else, these rotating rings only contain a series of random characters. If the colored data points are connected in the correct order, a large ethernet cable is plugged into what the players assume to be a window, and the players get downloaded into a virtual conference room. Due to the limitations discussed at the beginning of this section, only the first room with two learning goals was implemented.

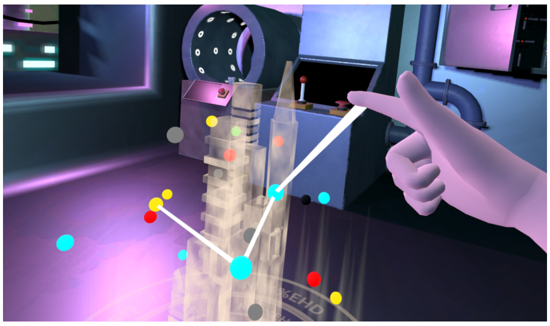


Figure 9: Figure illustrating the map table. The cryptographer can figure out the sequence by decrypting the code at the bottom of the map table, and the data analyst can trace the appropriate sequence by following the corresponding color codes.

3.4 Implementation

The project required 4 players to freely move around in a physical space whilst wearing VR gear providing an audiovisual virtual experience of the defined escape room. Additionally, in all stages of the puzzles, the players' avatars, as well as the objects the players interact with, had to be synchronised on all the other devices. This section elaborates on the tools that were used to create the prototype.

Regarding the technical details, the Unity engine [2] was utilised in conjunction with the C# programming language for the development of the VR prototype. Furthermore, PUN's Fusion [3] was leveraged for enhanced real-time physics networking capabilities and the availability of pre-hosted servers. These servers were ready to be used directly in the Unity editor right after modifying the networking scripts, which allowed for more efficient creation of synchronisation functionality. Remote Procedure Calls (RPCs) were extensively utilized to trigger synchronization among players, as well as networked objects and rigid bodies.

Concerning the visual design of the prototype, the Blender [4] software was exercised to create the 3D models and the visual environment for the VR experience. Moreover, Adobe's Substance Painter [5] was used to generate and modify textures for the materials of the 3D objects.

In terms of development, to ensure efficient collaboration and version control, Git [6] was employed via GitHub [7] for managing and tracking the prototype's codebase.

For audio effects, sound effects were sourced from Envato Elements [8], a resource for high-quality assets including audio. To tailor the audio to the requirements of the project, Audacity [9] was utilized, a versatile and free audio editing software.

The hardware setup consisted of four Meta Quest 2 headsets and their handheld controllers, which were borrowed from Aalborg University's Multisensory Experience Lab [10]. During the development phase,

[2] Unity engine: `https://unity.com/`
[3] Fusion networking: `https://www.photonengine.com/fusion`
[4] Blender: `https://www.blender.org/`
[5] Substance Painter: `https://substance3d.adobe.com/education`
[6] Git: `https://git-scm.com/`
[7] GitHub: `https://github.com/`
[8] Envato elements: `https://elements.envato.com/`
[9] Audacity: `https://www.audacityteam.org/`
[10] AAU ME Lab `https://melcph.create.aau.dk/`

a system was created that would allow runtime role assignment as well as debugging scripts. This setup permitted avoiding the need for constant physical space usage. Furthermore, the joysticks were utilized on the controllers for navigation within the virtual environments.

To represent players in the virtual environment, the Meta Avatars SDK [11] for Unity was employed, with the hopes to create a visually immersive experience. This SDK is, at the time of writing, natively supported by Meta's VR headsets, making it a particularly good choice based on the requirements and resources of this project.

Two tools were used to save time during the development phase. A Natural Language processing artificial intelligence called ChatGPT was used to generate random emails [12]. Furthermore, to facilitate realistic physics interactions within the virtual environment, the Hurricane VR [13] framework for Unity was employed.

It is worth mentioning that all the code and the prototype's history are available as a GitHub repository with the name "Did Stuff Escape", under the organization name "Did Stuff Studio" [14]. Given the extensive size and complexity of the codebase, this report restricts the inclusion of the actual code within the document. This approach was chosen to ensure that the report would maintain a concise presentation while providing readers with the opportunity to explore the code in its entirety.

4 Evaluation Methodology

To evaluate the prototype a quasi-experimental comparison group study design was used, with a non-probabilistic convenience sampling method. The independent variable in this between-group study was the type of compliance training administered, while the dependent variables were motivation, engagement, and perceived learning. The experimental group of this research received training through the VR educational ER. The test for this group was conducted in one of the seminar rooms of Aalborg University Copenhagen, where the test conductor gave participants an introduction, followed by the VR experience. After the experience participants were asked to fill out the questionnaires regarding the aforementioned dependent variables. Previously in the report, the time frame of the VR experience is mentioned to be 15 minutes. This is the same information that the participants were given during the testing, however, if the groups did not complete the experience in this amount of time they were allowed to finish it. This was done to ensure that measures of the dependent variable were taken from the same complete experience and to gather as much observational data regarding usability and playtesting as possible. The testing for the experimental group happened in groups of 4, while the members of the comparison group completed it individually.

The comparison group received the traditional training method of self-study e-learning. The e-learning material and activities were administered through Google Forms, followed by the same questionnaires as for the experimental group. Both groups' training was based on the same learning goals, highlighted in appendix B. The e-learning contained a brief written introduction, followed by a short description of the policy and the guidelines that apply to it. After the text, there were a few questions to aid the understanding of participants regarding the policy and its guidelines. There were a total of two sections like this, covering the same two learning goals as the VR training did. The first learning goal had three questions to promote comprehension, while the second had four. The e-learning training used in this evaluation can be seen in Appendix X. After the training portion, the participants had the same questionnaires regarding the dependent variables.

To measure the motivation of participants, the **Situational Motivational Scale (SIMS)** [21] was used. This scale aims to grasp the participants' motivational orientation towards a given activity. The SIMS is based on the same theory of self-

[11] Meta Avatars SDK: `https://developer.oculus.com/blog/meta-avatars-sdk-now-available/`

[12] ChatGPT: `https://chat.openai.com/`

[13] Hurricane VR: `https://cloudwalker2020.github.io/HurricaneVR-Docs/manual/intro.html`

[14] Codebase repository: `https://github.com/DidStuffStudio/Did-Stuff-Escape`

determination discussed earlier in this paper [Section 2.5]. Besides measuring intrinsic motivation, which is when engagement in an activity is for the satisfaction gained from it, it further divides extrinsic motivation into two different types and also adds a third concept of amotivation. **External regulation** is when part-taking in an activity is due to rewards or punishment, while identified regulation is performing an activity because it is perceived to be important or beneficial for the individual. Both of these are types of extrinsic motivation. **Amotivation** is neither extrinsic nor intrinsic, and is defined as the lack of drive to engage in any activity. When placed on a spectrum, intrinsic motivation is associated with the most positive outcomes, followed by identified regulation, external regulation, and finally amotivation at the other end associated with negative outcomes. Each of these four types of motivations has 4 items assigned to them, making the scale consist of 16 items in total. The items are statements that try to identify why an individual might be engaged in the activity. Each item has to be rated on a 1 - 7 Likert scale ranging from corresponds not at all to corresponds exactly. The resulting answers can be summed up and averaged per type of motivation for a measure that provides insight into why the individual is engaged in the given activity.

To second measure used in this study was the short form of **User Engagement Scale (UES-SF)** [35]. The UES-SF views engagement not only as a quality of user experience but also as the ability to engage in digital environments to achieve positive outcomes. This scale is a tool specifically developed to measure user engagement in a variety of digital environments (e.g. e-learning, video games) so that it can be used in design and evaluation. It comprises six dimensions. **Focused attention** refers to being intensely absorbed in an activity to the point of losing track of time, identical to what the theory of flow proposes. The dimension of **Perceived Usability** encompasses negative experiences stemming from the interactions, degree of control, and effort necessary to use a product. **Aesthetic appeal** is the attractive visual quality of an environment or interface. **Endurability** is the general success of the interaction and the users' inclination to return back to the product or endorse it to others. **Novelty** is if the interaction appealed to the curiosity and interest of the user, while felt **Involvement** is being captivated by the product. Each of these constituent factors has 3 items belonging to them, resulting in a total of 12 items. The items are statements regarding the previously listed dimensions. Every item has to be rated on a 5-point Likert-scale that ranges from 1 to 5, where 1 corresponds to strongly disagree and 5 corresponds to strongly agree. The items of perceived usability are negatively worded and as such they have to be reverse-coded before scoring. Items of the same dimension can be added together and averaged for a score that provides a more nuanced look into each factor of engagement, or all the items can be summed up and averaged for an overall engagement score.

The third measure used was the **CAP Perceived Learning Scale** [41]. This 9-item self-report scale was developed to evaluate the effectiveness of new educational opportunities and instructional technologies, such as 3D virtual worlds. The CAP measures perceived cognitive, affective, and psychomotor learning. It takes Bloom's Taxonomy [5] as the foundation for the definition of cognitive learning, which describes it as the ability to remember or identify knowledge and the expansion of intellectual capacities and competencies. The domain of affective learning is based on Kearney's explanation [27] which details it as the development of attitudes, behaviour, interests, views, values, and emotions rather than the cognitive faculties. An environment that nurtures affective learning leads to increased motivation, better engagement and thus positive learning outcomes [41]. The last domain of the scale is psychomotor learning, which refers to the development of skills related to physical motion [44]. It was decided that this last domain would not be included in the evaluation of this project, because the main focus of the research was more on the cognitive and affective outcomes. Without psychomotor learning, the CAP has only 3 items per domain, making it a total of 6 items. The statements in the scale have to be rated from a Likert-scale that ranges from 0 to 6, where 0 stands for not at all and 6 stands for very much so. One of the items within the perceived cognitive learning domain is negatively worded and has to be reverse-coded. Items can be summed up either per domain or all together to get a total CAP score. The higher the scores, the higher the perceived learning of individuals.

Additionally to the above-listed three measures, some extra questions were added to the survey. These questions aimed to inquire about the participants' perception of the ER in comparison with traditional methods they have tried before, their experiences with the team, if they would recommend it to their peers, and if they would like to see similar activities at their place of work, or school. These additional questions can be seen in appendix A.1.0.4. In total 9 extra questions were added, 7 of which were rated on a 5-point Likert scale, and the remaining two was a simple yes or no response. These questions were only added for the experimental group, and not for the comparison group.

5 Results

between group In total 32 people (n=32) participated in the study, between the ages of 19 to 37 ($\mu = 26.032, \sigma = 4.793$) out of which 20 were male (62.5%), 10 were female (31.25%), and 2 preferred not to specify (6.2%). From the total sample 12 people (37.5%) had prior experience with some form of compliance training. The 16 people (n_1=16) in the experimental group consisted of 12 men (75%) and 4 women (75%), between the ages of 19 to 37 ($\mu = 26.12, \sigma = 4.69$), with 5 people (31.2%) having prior experience in corporate compliance training. This group was made up of 4 teams of 4 people, out of which only 3 people (18.7%) had no prior experience with VR. None of the teams in the experimental group were able to complete the experience in the given amount of time. The average completion time was 33.57 minutes ($\sigma = 5.57$). The comparison group had 16 people (n_2=16) as well, consisting of 6 women (37.5%), 8 men (50%), and 2 (12.5%) who preferred not to say. Their ages ranged between 19 to 37 ($\mu = 26.12, \sigma = 4.95$) as well. The samples were drawn through convenience sampling. The experimental group's participants were students available at the University Campus, while the comparison group's participants were recruited from the researchers' acquaintances as well as from online forums dedicated to surveys.

Because of the small sample size and the sampling method used, the resulting scores from the applied measures were checked for statistically significant differences with the non-parametric Mann–Whitney U test. The scores and the calculations can be seen in appendix D.

In terms of motivation, statistically significant differences were identified in 3 out of the 4 subscales of the SIMS, as shown in table 3. The experimental group's intrinsic motivation ($\mu = 5.62, \sigma = 0.92$) and identified regulation ($\mu = 4.56, \sigma = 1.03$) were higher than the comparison group's ($\mu_{\mathrm{IM}} = 1.69, \mu_{\mathrm{IR}} = 1.69, \sigma_{\mathrm{IM}} = 1.23, \sigma_{\mathrm{IR}} = 1.14$). The external regulation scores of the experimental group ($\mu = 3.25, \sigma = 1.67$) were lower than the comparison group's ($\mu = 6.03, \sigma = 1.33$). There were no significant differences in the amotivation subscale.

SIMS	VR μ	VR σ	Traditional μ	Traditional σ
Intrinsic	5.62	0.92	1.69	1.23
Identified	4.56	1.03	2.41	1.14
External	3.25	1.67	6.03	1.33
Amotivation	2.91	1.01	3.97	1.83

Table 3: Comparison table between the means (μ) and standard deviations (σ) of the experimental group vs comparison group in the Situational Motivational Scale (SIMS).

In terms of engagement, both the total scores of the UES-SF and the individual dimensions were statistically different. The overall score of the experimental group ($\mu = 4.17, \sigma = 0.56$) was higher than the control group's ($\mu = 2.13, \sigma = 0.28$). The same can be said for all the constituent dimensions with the exception of perceived usability, which was lower for the group that received the VR training. The descriptive statistics for these dimensions for each group can be seen in table 4.

For perceived learning, the only statistically significant difference was on the affective subscale, where the experimental group ($\mu = 11.50, \sigma = 3.83$) scored higher than the group with the traditional e-

UES	VR μ	VR σ	Traditional μ	Traditional σ
Overall	4.17	0.56	2.13	0.28
FA	4.58	0.56	1.17	0.67
PU	3.62	0.72	4.73	1
AE	4.21	0.7	1.19	0.45
RW	4.27	0.72	1.44	1.08

Table 4: Comparison table between the means (μ) and standard deviations (σ) of the experimental group vs comparison group in the User Engagement Scale (UES).

learning method ($\mu = 6.81, \sigma = 3.62$). On the cognitive subscale and the total scores, no significant difference was detected. The descriptive statistics for these results can be seen in table 5.

CAP	VR μ	VR σ	Traditional μ	Traditional σ
Total	23.44	6.64	20.94	6.16
Cognitive	11.94	3.47	14.12	4.13
Affective	11.5	3.83	6.81	3.62

Table 5: Comparison table between the means (μ) and standard deviations (σ) of the experimental group vs comparison group in the CAP Perceived Learning scale.

The results of the additional questions administered to the experimental group are shown in table 6. Based on the results, it can be inferred that participants preferred the VR compared to traditional methods they have tried before, but they were not that confident that they have learned more than with e-learning self-study methods. The data also strongly indicates that participants of the VR Escape Room all felt like valuable parts of their team and had a collaborative experience. When asked if they would recommend this escape room to their peers 87.5% answered yes. When asked if they would like to see more activities like this at their place of work or school all participants said yes.

6 Discussion

Throughout the evaluation of the VR educational ER, a number of observations were made by the team. This section will begin with a discussion of these. During the introduction of the story that preceded the experience, participants all seemed to exhibit signs of curiosity and excitement. The occasional smiles, giggles, and other minor audible and physical reactions to the story being presented to them grew tenfold once they entered the virtual environment. Participants were visibly enjoying themselves, exploring their own and each other's avatars as well as the environment. Many of them proudly exclaimed which one of the 4 unique roles was assigned to them, which was followed by their team members joining in starting a group discussion about what their skills might be used for. In a sense, the in-game skills and the mystery of their purpose served as an icebreaker for teams to feel more relaxed with each other and the game's environment. For teams who are not that well acquainted, this is an excellent way to begin the experience.

An unexpected negative effect of the initial excitement from transferring into the VR experience, and getting roles assigned was forgetting the first clue given during the brief. It seemed like the amount of introductory information paired with a good deal of sensory stimuli from the foreign virtual environment resulted in a cognitive overload. To avoid this, the introduction should be shortened and given in the Virtual Environment through an element that players can access in case they need a refresher. Originally the plan was to have an introductory level, where players can get acquainted with the controls and the environment, while a brief is given to them through a non-player character (NPC). The design was however limited to only one level to decrease the complexity of the implementation, and therefore this tutorial level was scrapped. Through the testing, we found that if this in-game brief was included the NPC should repeat the introduction if prompted, or make key points in some other way available for participants.

Once the initial clue was recalled either through group discussion or a hint from the test conductor, participants quickly realized that the next step would involve the mail fetcher. It was a clear sign for players that the machines in the room needed electricity to function from their blank displays and lack of reactions when interacting with them. The physical puzzle in the circuit box also proved to be straightforward, including the fact that only the electrician role could operate it. The abilities of the cryptographer and data analyst were revealed to

Additional Questions	μ	σ
I liked this escape room better than traditional self-study online courses.	4.56	0.73
I learned more with this escape room than I would have with traditional self-study online courses.	3.88	1.15
I felt like I was part of the team.	4.69	0.48
I felt like I was contributing to the team.	4.5	0.89
My actions had an impact on the team.	4.5	0.89
The actions of other team members affected me.	4.62	0.62
I felt left out.	1.75	0.93

Table 6: Mean (μ) and standard deviation (σ) of the additional questions asked to the experimental group.

the teams through effective communication. Teams that vocalized their experiences were faster in figuring out that some players could see or read things that others could not.

The scanning of the e-mail as a way to report it was somewhat unclear for players, which could also be explained by the initial cognitive overload. Another explanation is that this element itself was adding to the cognitive burden by containing references to multiple ethics violations and also clues for the next puzzle. A possible way to tackle this issue would be to reduce the information in the e-mail by only including one ethics violation. The next clue should be made available only after scanning the mail and categorizing the violation through the flagging system. The cognitive load could also be reduced simply by rewording the contents of the display above the window from "0 out of two violations detected" to "0 out of two violations scanned". This display proved useful when informing players that new elements of the environment were unlocked.

Identifying violations through scanning and categorizing them was the main method of delivering learning goals. The aim of this activity was to allow players to analyze, evaluate and discuss the information they are presented with by the flagging system regarding the ethics breach found in the clue. In essence, it was a form of scenario-based learning to support the active learning process and induce higher-order thinking. Teams intuitively recognized elements that contained violation(s) and after scanning discussed what parts of the presented information pertains to them. The voting system in this section successfully encouraged players to share their perspectives and listen to others. When the second violation was found in the room, some players seemed frustrated by the fact that they had to repeat the same task again and that it involved a lot of text. A way to avoid this frustration could be to change up the method of categorization for each violation. Instead of doing the same thing over and over, players should be presented with new novel mini-games that still support higher-order cognitive processes and collaboration.

Another element in the game that was somewhat unsuccessful was the machine to create objects. The naming, the function, and the involved mini-puzzle all seemed too foreign to participants. It was also unclear that this section was specifically designed for the skills of the programmer. It could be that the frustration of not understand the function and not being able to operate it effectively was one of the reasons contributing to the unsatisfactory results on the perceived usability subscale of the UES-SF.

Based on the observations and the scores of the CAP perceived learning scale, it is no wonder that no statistically significant cognitive learning was perceived when such a heavy burden was placed on the participant's mental faculties by other elements of the game. As shown by the means in table 5, traditional e-learning scored higher than the VR experience in this subscale. The cognitive overload that prevented learning could also have been further aggravated by other usability issues, indicated by the poor perceived usability scores from the UES-SF. The affective perceived learning scores of the escape room, however, were statistically different, demonstrating that some learning in terms of attitudes, values, behaviours, or emotions was perceived. Considering the topic of the training, this is a good implication that it had a positive impact on learning outcomes. It indicates that the scenario-based flagging system paired with the immersiveness of the experience has potential in improving compliance training. To get more concrete evidence that supports this claim, comparing the VR experience to other forms of scenario-based training methods such as facilitated training with case-studies could be done.

In terms of motivation, the results showed that the VR experience stimulated intrinsic motivation more than its traditional counterpart. This indicates a positive and enjoyable experience, which is further supported by the results of the engagement questionnaire [Table 4] and the additional questions [Table 6], which can be seen in appendix D. Participants of the experimental group also seemed to identify their experience as more personally beneficial and much less externally motivated. The engagement measure revealed that the VR training was an aesthetically pleasing and absorbing experience that appealed to their curiosity and captivated their interest. The interaction as a whole might have been rewarding and could be called a success, but the experience has usability issues that need to be further explored and tended to. Further usability and playtesting is necessary to polish the experience and maximize its effect. The immersiveness of the experience somewhat clashed with the escape room format. Because participants were drawn in to the activity, their time perception might have been distorted. Teams often times did not seem like they were in a hurry to "escape". To create more sense of urgency, a countdown timer should be added to the experience.

One notable usability concern arose from the inconsistent availability of internet connectivity. In cases where a headset encountered temporary disconnection, users were able to rejoin the session. However, this led to disruptions in assigned roles and a failure to synchronize the state of the virtual room, such as the operational status of machines.

This design limitation necessitated a complete reset of the experience, resulting in a disruption of narrative continuity. Although contingency protocols were implemented to regenerate certain primary story objects and facilitate a partial resumption of the experience, it became evident that this aspect generated significant frustration among participants. To address this issue, a potential solution would involve the implementation of a system that automatically synchronizes the state of the virtual room upon rejoining, while actively searching for any unfulfilled roles.

Furthermore, the existing flagging system suffered from a deficiency in design, as it failed to deliver information in a user-friendly manner. Users were confronted with an overwhelming influx of text and an array of decision-making options, which could be considered intimidating. Improving the system's effectiveness would have been possible through an iterative playtesting process, exploring different methods of flagging ethics violations. For example, the inclusion of a virtual ballot box or a tablet-like interface could have been considered as a potential solution.

The sample of the testing consisted of individuals under 40, some of whom had jobs where compliance training is relevant. To be able to make more robust conclusions about this experience, it would be ideal to test it with a company in a case-study format. To ensure good usability for the whole target audience participants from the 40 to 65 age group should also be involved in future evaluations.

7 Conclusion

This research paper aimed to investigate whether virtual reality escape rooms developed for corporate compliance training are a better alternative to traditional training methods. Relevant theories about active learning, escape rooms for education, game-based learning, motivation, immersion, and flow were investigated to serve as a guide for designing the experience. The learning goals for the VR training were based on publicly available policies of large corporations.

To evaluate the VR experience that was developed based on the requirements synthesized from the analysis, a quasi-experimental comparison study measuring engagement, motivation, and perceived learning of participants was used. The experimental group received the VR escape room compliance training, while the comparison group received the popular and widely-adopted traditional e-learning self-study training. In terms of engagement and motivation, the scores of the VR experience were statistically significantly better than the traditional training. Therefore the following null hypotheses can be rejected:

$H0_a$: Virtual reality escape rooms for compliance training have no effect on the engagement of employees compared to the traditional e-learning self-study method.

> **$H0_b$**: Virtual reality escape rooms for compliance training have no effect on the motivation of employees compared to the traditional e-learning self-study method.

In the domain of perceived learning, there was a significant difference in the affective subscale, with the VR having a more positive impact on learning. Because no significant difference was detected between the cognitive subscale and the total perceived learning score, the following null hypothesis can not be rejected:

> **$H0_c$**: Virtual reality escape rooms for compliance training have no effect on the perceived learning of employees compared to the traditional e-learning self-study method.

The above three hypotheses when combined together result in a statement that posits that engagement, motivation, and perceived learning all remain unaffected by VR training compared to the traditional method. The statement is as follows:

> ***H0**: Virtual reality escape rooms for compliance training have no effect on the engagement, motivation, and perceived learning of employees compared to the traditional e-learning self-study method.*

This combined null hypothesis can not be completely refuted. Significant changes were detected in two out of the three measures, and also in a subcategory of the third measure. Taking into consideration all of the above-mentioned results, there is a strong indication that further development and refining of the VR experience has the potential to fully reject the null hypothesis (H0).

Participants' initial excitement and curiosity were evident, indicating that the introduction of unique roles effectively served as an icebreaker for the teams. This could be especially beneficial for groups that are less familiar with each other. However, the cognitive overload experienced during the introductory stage resulted in some participants forgetting the first clue. This suggests that adjustments need to be made in the delivery of the introductory information, perhaps through the inclusion of an in-game brief for refreshers.

The game mechanisms, which allowed players to utilize their unique abilities, effectively encouraged teamwork and communication. However, certain elements like the e-mail scanning system and the machine to create objects were challenging for participants. These findings point to areas where the experience could be improved.

Despite the cognitive load, the VR experience showed potential in delivering learning goals through scenario-based learning. The voting system promoted collaboration and communication among participants, suggesting that the immersive, collaborative aspects of VR could be beneficial for learning environments.

Interestingly, while the VR experience did not result in a significant perceived cognitive learning outcome when compared to traditional e-learning, it performed better in terms of affective perceived learning. This indicates that the immersive experience could have a positive impact on attitudes, behaviors, and emotions, even if the cognitive learning outcomes are similar to traditional methods.

The VR experience significantly stimulated intrinsic motivation and was reported as an enjoyable experience, despite some usability issues. These findings highlight the need for further testing and adjustments to maximize the effectiveness of the VR experience.

Major sources of frustration among participants included internet connectivity issues and the lack of an automated system to synchronize the state of the virtual room upon reconnection. These findings underscore the need for technical improvements to address these issues.

In terms of the flagging system, its overly complicated nature due to the sheer amount of text and decision-making options presented to the users indicates a need for future design iterations to explore more user-friendly alternatives.

Lastly, while the testing was performed with a sample group under 40, future evaluations should consider including participants from the 40 to 65 age group. This would ensure a comprehensive understanding of the VR experience's effectiveness across the entire target audience, paving the way for more robust conclusions and potentially improving the experience's usability and effectiveness for a wider range of users.

References

[1] Nets Denmark A/S. *Nets Ethical Group Policy*. https://www.nets.eu/about-nets/csr/Documents/Nets-Ethical-Policy.pdf. Accessed: May 25, 2023. 2022.

[2] Azita Iliya Abdul Jabbar and Patrick Felicia. "Gameplay engagement and learning in game-based learning: A systematic review". In: *Review of educational research* 85.4 (2015), pp. 740–779.

[3] Avshalom M Adam and Dalia Rachman-Moore. "The methods used to implement an ethical code of conduct and employee attitudes". In: *Journal of Business Ethics* 54 (2004), pp. 225–244.

[4] TalentLMS Aris Apostolopoulos. *The 2019 Gamification at Work Survey*. https://www.talentlms.com/blog/gamification-survey-results/. Accessed: May 25, 2023. 2019.

[5] Benjamin S Bloom et al. "Taxonomy of educational objectives: Handbook I". In: *Cognitive domain. New York: David McKay* (1956).

[6] Charles C Bonwell and James A Eison. *Active learning: Creating excitement in the classroom. 1991 ASHE-ERIC higher education reports.* ERIC, 1991.

[7] Cynthia Brame. "Active learning". In: *Vanderbilt University Center for Teaching* (2016).

[8] Emily Brown and Paul Cairns. "A grounded investigation of game immersion". In: *CHI'04 extended abstracts on Human factors in computing systems*. 2004, pp. 1297–1300.

[9] Paul Cairns, Anna Cox, and A Imran Nordin. "Immersion in digital games: review of gaming experience research". In: *Handbook of digital games* (2014), pp. 337–361.

[10] Elucidat Charlotte Curl. *10 revealing compliance training research stats for financial services [Infographic]*. https://www.elucidat.com/blog/research-compliance-training/. Accessed: May 25, 2023. 2020.

[11] Luca Chittaro and Fabio Buttussi. "Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety". In: *IEEE transactions on visualization and computer graphics* 21.4 (2015), pp. 529–538.

[12] Helen Colman. *Learning Retention – How To Make Information Stick*. https://www.ispringsolutions.com/blog/learning-retention. Accessed: May 25, 2023. 2022.

[13] Sara De Freitas. "Learning in immersive worlds: A review of game-based learning". In: (2006).

[14] Edward L. Deci et al. "Motivation and Education: The Self-Determination Perspective". In: *Educational Psychologist* 26.3-4 (1991), pp. 325–346. DOI: 10.1080/00461520.1991.9653137. eprint: https://doi.org/10.1080/00461520.1991.9653137. URL: https://doi.org/10.1080/00461520.1991.9653137.

[15] StatBank Denmark. *Statistics Denmark*. https://www.statbank.dk/. Accessed: May 25, 2023. 2023.

[16] Heather Desurvire and Max Kreminski. "Are game design and user research guidelines specific to virtual reality effective in creating a more optimal player experience? Yes, VR PLAY". In: *Design, User Experience, and Usability: Theory and Practice: 7th International Conference, DUXU 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part I 7*. Springer. 2018, pp. 40–59.

[17] Virtual Speech Dom Barnard. *History of VR - Timeline of Events and Tech Development*. https://virtualspeech.com/blog/history-of-vr. Accessed: May 25, 2023. 2022.

[18] Inc. — XR Training Elite XR. *XR for Oil and Gas*. https://elitevrtraining.com/oil-and-gas/. Accessed: May 25, 2023. 2023.

[19] Richard M Felder and Rebecca Brent. "Active learning: An introduction". In: *ASQ higher education brief* 2.4 (2009), pp. 1–5.

[20] T Fulya Eyupoglu and John L Nietfeld. "Intrinsic motivation in game-based learning environments". In: *Game-based assessment revisited* (2019), pp. 85–102.

[21] Frédéric Guay, Robert J Vallerand, and Céline Blanchard. "On the assessment of situ-

ational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS)". In: *Motivation and emotion* 24 (2000), pp. 175–213.

[22] Juho Hamari, Jonna Koivisto, and Harri Sarsa. "Does Gamification Work? – A Literature Review of Empirical Studies on Gamification". In: *2014 47th Hawaii International Conference on System Sciences*. 2014, pp. 3025–3034. DOI: 10.1109/HICSS.2014.377.

[23] HSI. *Compliance Training Topics*. https://hsi.com/solutions/employee-training-and-development/compliance-training-topics. Accessed: May 25, 2023.

[24] Hui-Ching Kayla Hsu, Cong Vivi Wang, and Chantal Levesque-Bristol. "Reexamining the impact of self-determination theory on learning outcomes in the online learning environment". In: *Education and information technologies* 24 (2019), pp. 2159–2174.

[25] Charlene Jennett et al. "Measuring and defining the experience of immersion in games". In: *International Journal of Human-Computer Studies* 66.9 (2008), pp. 641–661. ISSN: 1071-5819. DOI: https://doi.org/10.1016/j.ijhcs.2008.04.004. URL: https://www.sciencedirect.com/science/article/pii/S1071581908000499.

[26] Muel Kaptein. "The effectiveness of ethics programs: The role of scope, composition, and sequence". In: *Journal of Business Ethics* 132 (2015), pp. 415–431.

[27] P Kearney. "Affective learning". In: *Communication research measures: A sourcebook* (1994), pp. 81–85.

[28] SARIA SE & Co. KG. *The SARIA Code of Conduct*. https://www.daka.dk/wp-content/uploads/2023/05/SARIA_Code_of_conduct_230227_EN.pdf. Accessed: May 25, 2023. 2023.

[29] Sonsoles López-Pernas et al. "Examining the use of an educational escape room for teaching programming in a higher education setting". In: *IEEE Access* 7 (2019), pp. 31723–31737.

[30] Celia Moore and Francesca Gino. "Ethically adrift: How others pull our moral compass from true North, and how we can fix it". In: *Research in Organizational Behavior* 33 (2013), pp. 53–77. ISSN: 0191-3085. DOI: https://doi.org/10.1016/j.riob.2013.08.001. URL: https://www.sciencedirect.com/science/article/pii/S0191308513000026.

[31] EQS Group Moritz Homann. *Compliance Officers – Duties, Qualifications and Remuneration*. https://www.eqs.com/compliance-blog/compliance-officer/#:~:text=Compliance%20Officers%20are%20responsible%20for%20ensuring%20that%20all%20corporate%20processes,comply%20with%20internal%20standards%20too. Accessed: May 25, 2023. 2023.

[32] Jeanne Nakamura, Mihaly Csikszentmihalyi, et al. "The concept of flow". In: *Handbook of positive psychology* 89 (2002), p. 105.

[33] S Nicholson. *Ask why: Creating a better player experience through environmental storytelling and consistency in escape room design (Vols. 1–17). Meaningful Play 2016*. 2016.

[34] Scott Nicholson. "Peeking behind the locked door: A survey of escape room facilities". In: (2015).

[35] Heather L. O'Brien, Paul Cairns, and Mark Hall. "A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form". In: *International Journal of Human-Computer Studies* 112 (2018), pp. 28–39. ISSN: 1071-5819. DOI: https://doi.org/10.1016/j.ijhcs.2018.01.004. URL: https://www.sciencedirect.com/science/article/pii/S1071581918300041.

[36] Oculus. *Oculus Quest Health & Safety Warnings*. https://securecdn.oculus.com/sr/oculusquest-warning-english. Accessed: May 25, 2023.

[37] Oersted Policy. *Oersted Good business conduct*. https://orsted.com/-/media/www/docs/corp/com/sustainability/good-business-conduct-uk-dk.pdf. Accessed: May 25, 2023. 2019.

[38] PowerDMS. *What is corporate compliance, and why is it important?* https://www.powerdms.com/policy-learning-center/what-corporate-compliance-is-and-why-its-important. Accessed: May 25, 2023. 2020.

[39] Thomson Reuters. *Building a compliance department.* https://legal.thomsonreuters.com/en/insights/articles/building-a-compliance-department. Accessed: May 25, 2023. 2021.

[40] Strategic Management Services Richard P. Kusserow. *Compliance Training Best Practices & Methods.* https://www.compliance.com/resources/effective-compliance-training-methods/. Accessed: May 25, 2023. 2020.

[41] Alfred P. Rovai et al. "Development of an instrument to measure perceived cognitive, affective, and psychomotor learning in traditional and virtual classroom higher education settings". In: *The Internet and Higher Education* 12.1 (2009), pp. 7–13. ISSN: 1096-7516. DOI: https://doi.org/10.1016/j.iheduc.2008.10.002. URL: https://www.sciencedirect.com/science/article/pii/S109675160800064X.

[42] Eric Sanchez and Maud Plumettaz-Sieber. "Teaching and learning with escape games from debriefing to institutionalization of knowledge". In: *Games and Learning Alliance: 7th International Conference, GALA 2018, Palermo, Italy, December 5–7, 2018, Proceedings 7.* Springer. 2019, pp. 242–253.

[43] Yen-Ru Shi and Ju-Ling Shih. "Game factors and game-based learning design model". In: *International Journal of Computer Games Technology* 2015 (2015), pp. 11–11.

[44] Elizabeth J Simpson. "The classification of educational objectives, psychomotor domain." In: (1966).

[45] Eugene Soltes and Hui Chen. *Why Compliance Programs Fail - and How to Fix Them.* https://edisciplinas.usp.br/pluginfile.php/4247829/mod_resource/content/0/SOLTES%2C%20Eugene%3B%20CHEN%2C%20Hui.%20Why%20compliance%20programs%20fail%20-%20and%20how%20to%20fix%20then.%20Harvard%20Business%20Review%2C%20march.april%202018..pdf. Accessed: May 25, 2023. 2018.

[46] MAN Energy Solutions. *MAN Energy Solutions Code of Conduct.* https://www.man-es.com/docs/default-source/compliance/man-energy-solutions-code-of-conduct_en.pdf?sfvrsn=fa495111_58. Accessed: May 25, 2023. 2021.

[47] Thomas Stöber, Peter Kotzian, and Barbara E Weißenberger. "Design matters: on the impact of compliance program design on corporate ethics". In: *Business research* 12 (2019), pp. 383–424.

[48] Alice Udvari-Solner. "Collaborative Learning". In: *Encyclopedia of the Sciences of Learning.* Ed. by Norbert M. Seel. Boston, MA: Springer US, 2012, pp. 631–634. ISBN: 978-1-4419-1428-6. DOI: 10.1007/978-1-4419-1428-6_817. URL: https://doi.org/10.1007/978-1-4419-1428-6_817.

[49] Neil Vaughan et al. "A review of virtual reality based training simulators for orthopaedic surgery". In: *Medical Engineering & Physics* 38.2 (2016), pp. 59–71. ISSN: 1350-4533. DOI: https://doi.org/10.1016/j.medengphy.2015.11.021. URL: https://www.sciencedirect.com/science/article/pii/S1350453315002799.

[50] Alice Veldkamp et al. "Escape education: A systematic review on escape rooms in education". In: *Educational Research Review* 31 (2020), p. 100364. ISSN: 1747-938X. DOI: https://doi.org/10.1016/j.edurev.2020.100364. URL: https://www.sciencedirect.com/science/article/pii/S1747938X20300531.

[51] Lev Semenovich Vygotsky and Michael Cole. *Mind in society: Development of higher psychological processes.* Harvard university press, 1978.

[52] Markus Wiemker, Errol Elumir, and Adam Clare. "Escape room games". In: *Game based learning* 55 (2015), pp. 55–75.

[53] Zhonggen Yu, Mingle Gao, and Lifei Wang. "The effect of educational games on learning outcomes, student motivation, engagement and satisfaction". In: *Journal of Educational Computing Research* 59.3 (2021), pp. 522–546.

[54] Luisa Zhou. *eLearning Statistics: The Definitive List in 2023*. https://www.luisazhou.com/blog/elearning-statistics/. Accessed: May 25, 2023. 2023.

[55] Olha Zubarieva. *A Comprehensive Guide to the Ideal Online Course Length*. https://sendpulse.com/blog/online-course-length#:~:text=However%2C%20if%20you're%20considering,t%20affect%20the%20student's%20engagement.&text=However%2C%20there%20is%20also%20a,between%2015%20and%2030%20minutes. Accessed: May 25, 2023. 2022.

Appendices

A Questions used during testing

A.1 VR escape room questions

This section outlines the questions that the participants of the experimental group had to answer after participating in the testing of the VR escape room prototype. They are listed as follows.

A.1.0.1 Situational Motivational Scale (SIMS).

1. Because I think that this activity is interesting.
2. Because I am doing it for my own good.
3. Because I am supposed to do it.
4. There may be good reasons to do this activity, but personally I don't see any.
5. Because I think that this activity is pleasant.
6. Because I think that this activity is good for me.
7. Because it is something that I have to do.
8. I do this activity but I am not sure if it is worth it.
9. Because this activity is fun.
10. By personal decision.
11. Because I don't have any choice.
12. I don't know; I don't see what this activity brings me.
13. Because I feel good when doing this activity.
14. Because I believe that this activity is important for me.
15. Because I feel that I have to do it.
16. I do this activity, but I am not sure it is a good thing to pursue it.

The following codification keys refer to the subscales of SIMS:

- Intrinsic motivation: Items 1, 5, 9, 13;
- Identified regulation: Items 2, 6, 10, 14;
- External regulation: Items 3,7, 11, 15;
- Amotivation: Items 4, 8, 12, 16;

A.1.0.2 User Engagement Scale Short Form (UES-SF).

The following list also specifies the subscale the question refers to. Focused attention (FA), Aesthetic Appeal (AE), Perceived Usability (PU), Reward Factor (RW):

1. FA-S.1: I lost myself in this experience.
2. FA-S.2: The time I spent using this escape room just slipped away.

3. FA-S.3: I was absorbed in this experience.
4. PU-S.1: I felt frustrated while using this escape room.
5. PU-S.2: I found this escape room confusing to use.
6. PU-S.3: Using this escape room was taxing.
7. AE-S.1: The escape room was attractive.
8. AE-S.2: The escape room was aesthetically appealing.
9. AE-S.3: The escape room appealed to my senses.
10. RW-S.1: Using the escape room was worthwhile.
11. RW-S.2: My experience was rewarding.
12. RW-S.3: I felt interested in this experience.

A.1.0.3 CAP Perceived Learning Scale.

This list excludes the psychomotor subscale.

- 1: I can organize the training material into a logical structure.
- 2: I cannot produce a training study guide for future students.
- 4: I have changed my attitude about the training subject matter as a result of this training.
- 5: I can intelligently critique the material used in this training.
- 6: I feel more self-reliant as a result of the content learned in this training.
- 9: I feel that I am a more sophisticated thinker as a result of this training.

The following codification keys refer to the subscales of CAP Perceived Learning Scale:

- Cognitive subscale: items 1, 2, and 5.
- Affective subscale: items 4, 6, and 9.
- Psychomotor subscale: items 3, 7, and 8 (excluded from this study).

A.1.0.4 Extra questions.

1. I liked this escape room better than traditional self-study online courses.
2. I learned more with this escape room than I would have with traditional self-study online courses.
3. I felt like I was part of the team.
4. I felt like I was contributing to the team.
5. My actions had an impact on the team.
6. The actions of other team members affected me.
7. I felt left out.
8. Would you recommend other students or colleagues to participate in the escape room? (Yes/No)
9. Would you like other courses (work or school) to include activities like this? (Yes/No)

A.2 Traditional compliance training questions

This section outlines the questions that the comparison group had to answer regarding traditional methods of ethical compliance training in corporations. They are listed as follows.

A.2.0.1 Situational Motivational Scale (SIMS).

1. Because I think that this activity is interesting.
2. Because I am doing it for my own good.
3. Because I am supposed to do it.
4. There may be good reasons to do this activity, but personally I don't see any.
5. Because I think that this activity is pleasant.
6. Because I think that this activity is good for me.
7. Because it is something that I have to do.
8. I do this activity but I am not sure if it is worth it.
9. Because this activity is fun.
10. By personal decision.
11. Because I don't have any choice.
12. I don't know; I don't see what this activity brings me.
13. Because I feel good when doing this activity.
14. Because I believe that this activity is important for me.
15. Because I feel that I have to do it.
16. I do this activity, but I am not sure it is a good thing to pursue it.

The following codification keys refer to the subscales of SIMS:

- Intrinsic motivation: Items 1, 5, 9, 13;
- Identified regulation: Items 2, 6, 10, 14;
- External regulation: Items 3,7, 11, 15;
- Amotivation: Items 4, 8, 12, 16;

A.2.0.2 User Engagement Scale Short Form (UES-SF).

The following list also specifies the subscale the question refers to. Focused attention (FA), Aesthetic Appeal (AE), Perceived Usability (PU), Reward Factor (RW):

1. FA-S.1: I lost myself in this experience.
2. FA-S.2: The time I spent using this online training just slipped away.
3. FA-S.3: I was absorbed in this experience.
4. PU-S.1: I felt frustrated while using this online training.
5. PU-S.2: I found this online training confusing to use.

6. PU-S.3: Using this online training was taxing.
7. AE-S.1: The online training was attractive.
8. AE-S.2: The online training was aesthetically appealing.
9. AE-S.3: The online training appealed to my senses.
10. RW-S.1: Using the online training was worthwhile.
11. RW-S.2: My experience was rewarding.
12. RW-S.3: I felt interested in this experience.

A.2.0.3 CAP Perceived Learning Scale.

This list excludes the psychomotor subscale.

- 1: I can organize the training material into a logical structure.
- 2: I cannot produce a training study guide for future students.
- 4: I have changed my attitude about the training subject matter as a result of this training.
- 5: I can intelligently critique the material used in this training.
- 6: I feel more self-reliant as a result of the content learned in this training.
- 9: I feel that I am a more sophisticated thinker as a result of this training.

The following codification keys refer to the subscales of CAP Perceived Learning Scale:

- Cognitive subscale: items 1, 2, and 5.
- Affective subscale: items 4, 6, and 9.
- Psychomotor subscale: items 3, 7, and 8 (excluded from this study).

B Learning Goals

The following list enumerates all the learning goals that were aimed to be fulfilled with the use of the VR escape room. In the actual prototype, due to the nature of the chosen narrative for the escape room, only combating corruption (1) and promoting human rights (6) were used.

1. Understanding and identifying different types of **corruption** such as abuse of power, nepotism, or bribery, and how to report them.
2. Understanding the company's policy on gifts and hospitality and how to assess their reasonableness, appropriateness, and justifiability, and how to report doubts or concerns.
3. Understanding and identifying situations that may pose a conflict of interest, and the importance of disclosing them.
4. Understanding the company's GDPR program and guidelines for safeguarding personal, private, and confidential information, and how to report discrepancies.
5. Understanding and complying with competition law, including not discussing confidential commercial matters with competitors and obtaining preapproval from Legal for approved sales templates.
6. Understanding and **promoting human rights**, including treating colleagues with respect, speaking up against discrimination, harassment, retaliation, and human rights abuses, and assessing and engaging with suppliers.
7. Understanding and complying with guidelines for engaging in political activities, including making economic contributions or donations to political parties, coordinating dialogues through Public Affairs, and registering in the EU Transparency Register.
8. Understanding and complying with guidelines for creating value through sponsorships and donations, including aligning with corporate values and business objectives, obtaining approval from the immediate manager, and coordinating with Public Affairs and CSR.
9. Understanding and adhering to guidelines for professional communication, including demonstrating professionalism in interactions with customers, suppliers, and business partners, not sharing confidential information on social media, and following social media guidelines.
10. Understanding and applying guidelines for ethical decision-making in social dilemmas, including considering legality, fairness, and personal discomfort.

C Narrative

The following text was read aloud to the experiment group as an introduction to the narrative of the VR escape room.

> Welcome! We have gathered you all here today because you are this company's most trusted and skilled experts. Our new Artificial Intelligence called Ethics Violation Authority, EVA for short, has detected some serious violations in our company. EVA traced back the violations starting from an Anonymous e-mail that was sent to Hanna Johnson at HR in November 2053. Unfortunately, before EVA could provide us with the details, someone hacked into her mainframe and started shutting her down. We are currently locked out of the system, and cannot fix her with ordinary methods, nor can we access any of the data regarding the ethics violations she has detected.
>
> Thanks to the company's latest top-secret technology, we can upload your consciousnesses to the cyber world, where in theory, you will be able to trace the ethics violations. Since the hacker is currently active in our system, we estimate that all of our data will be deleted in about 15 minutes, so time is of the essence. Before shutting down, EVA told us that she found in total 2 major violations and multiple minor ones. Your job will be to find, scan, flag, and confirm these violations. We will place your uploaded consciousness into the e-mail directory of the company's network, but we don't exactly know what you will be faced with. The technology is entirely untested, but we know that you will have different capabilities or skills in the cyber world depending on the structure of your uploaded consciousness.
>
> Amongst you, there will be:
>
> - An electrician – specializes in handling electricity in the cyberworld, such as powering machinery.
> - A programmer – can use computers inside the cyberworld to fabricate objects with ease.
> - A cryptologist – is able to read encrypted messages.
> - A Data Analyst – will be able to see the flow of data in some cases.

D Calculation of UES, CAP and SIMS scores

This Jupyter Notebook performs an evaluation using User Engagement Scale Short Form (UES - SF), CAP perceived learning, and Situational Motivational Scale (SIMS). The resulting questionnaire data from the tests was compiled into CSV format: one for the experimental group that tried the VR escape room, and another one for the comparison group regarding traditional compliance training in corporations.

The notebook calculates scores for each scale and provides insights into the participants' experiences and perceptions.It also includes additional statistical analyses such as the Mann-Whitney U test for comparing scores between different groups, as well as means and standard deviations for each subscale. The calculated scores and statistical tests provide valuable insights into participants' experiences, perceptions, and learning outcomes.

In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import scipy.stats as stats
from tabulate import tabulate
from scipy.stats import mannwhitneyu
```

The following code snippet reads data from two CSV files, 'Testing_EscapeRoomVR.csv' and 'TraditionalComplianceTrainingTesting.csv', into pandas DataFrames. It also removes leading and trailing spaces from the column names of the DataFrames to ensure consistency and cleanliness of the data.

In [ ]:
```
# Read the CSV file
data_vr = pd.read_csv('Testing_EscapeRoomVR.csv')
data_traditional = pd.read_csv('TraditionalComplianceTrainingTesting.csv')

# Remove leading and trailing spaces from column names
data_vr.columns = data_vr.columns.str.strip()
data_traditional.columns = data_traditional.columns.str.strip()
```

The following code defines a function called `PerformMannWhitneyUTest` that takes two samples as input and performs the Mann-Whitney U test.

The **Mann-Whitney U test** compares the ranks of the observations between the two samples and assesses whether one sample tends to have larger values than the other. The test evaluates the following hypotheses:

- Null Hypothesis (H0): The distributions of both samples are equal.
- Alternative Hypothesis (H1): The distributions of the two samples are significantly different.

If the calculated p-value is less than the significance level (p-value < α), it suggests that there is sufficient evidence to reject the null hypothesis, indicating a significant difference between the two samples.

In [ ]:
```
def PerformMannWhitneyUTest(sample1, sample2):
    alpha = 0.05

    # Perform Mann-Whitney U test
    statistic, p_value = mannwhitneyu(sample1, sample2)

    # Print the results
    print("Statistic:", statistic)
    print("p-value:", p_value)
    print("Mann Whiney U Test --> Are the samples significantly different? --> ", p_value < alp
```

The `CalculateStats` function calculates the mean and standard deviation of a given set of scores and prints the results. It takes in the scores and a title as input and provides a summary of the statistical properties of the scores. The `PlotScoresBoxPlot` function simply draws boxplots for scores.

In [ ]:
```
def CalculateStats(scores, title):
    # Calculate mean and standard deviation
    mean = scores.mean()
    std = scores.std()

    # Print mean and standard deviation title and description in same string
    print(f"{title} --> Mean: {mean:.2f}, Standard Deviation: {std:.2f}")
    print("\n\n")

    return mean, std
```

```
# plot the scores of subscales. the subscale data parameter should include the overall score as
def PlotScoresBoxPlot(subscale_titles, subscale_data, description = "Boxplots of Scores"):
    # Generate boxplots for overall score + subscale scores in a single figure
    fig, axes = plt.subplots(nrows=1, ncols= len(subscale_data), figsize=(10, 6))

    for i, ax in enumerate(axes.flat):
        ax.boxplot(subscale_data[i])
        ax.set_title(subscale_titles[i])
        ax.set_ylabel('Score')

    # Add a common title for the entire figure
    fig.suptitle(description)

    # Adjust layout and display the figure
    fig.tight_layout()
    plt.show()
```

Calculation of UES Scores

This function takes input data and a dictionary of UES items as parameters. It calculates the overall engagement score and subscale scores based on the UES items. Following the scoring instructions of UES-SF, the following function reverses the scores for specific items (PU-S), calculates the overall participant scores as well as subscale scores. Finally, it displays the results in a table format using the tabulate library. The function returns the participant scores and subscale scores as output.

In [ ]:

```
def CalculateScoreUES(data, ues_items, description):
    ues_columns = list(ues_items.values())

    # Subset the data with UES columns
    ues_data = data[ues_columns].copy()

    # Reverse code the items PU-S.1, PU-S.2, PU-S.3
    ues_data[ues_items['PU-S.1']] = 6 - ues_data[ues_items['PU-S.1']]
    ues_data[ues_items['PU-S.2']] = 6 - ues_data[ues_items['PU-S.2']]
    ues_data[ues_items['PU-S.3']] = 6 - ues_data[ues_items['PU-S.3']]

    # Calculate the overall UES-SF score per participant
    ues_participant_scores = ues_data.mean(axis=1)

    # Calculate subscale scores per participant
    ues_subscale_scores = {
        'FA-S': ues_data[[ues_items['FA-S.1'], ues_items['FA-S.2'], ues_items['FA-S.3']]].mean(
        'PU-S': ues_data[[ues_items['PU-S.1'], ues_items['PU-S.2'], ues_items['PU-S.3']]].mean(
        'AE-S': ues_data[[ues_items['AE-S.1'], ues_items['AE-S.2'], ues_items['AE-S.3']]].mean(
        'RW-S': ues_data[[ues_items['RW-S.1'], ues_items['RW-S.2'], ues_items['RW-S.3']]].mean(
    }

    # Prepare the data for the table
    table_data = [
        ['Participant', 'Overall Engagement Score', 'FA-S Score', 'PU-S Score', 'AE-S Score', '
    ]

    for i, score in enumerate(ues_participant_scores):
        fa_s_score = ues_subscale_scores['FA-S'].iloc[i]
        pu_s_score = ues_subscale_scores['PU-S'].iloc[i]
        ae_s_score = ues_subscale_scores['AE-S'].iloc[i]
        rw_s_score = ues_subscale_scores['RW-S'].iloc[i]
        table_data.append([f'Participant {i+1}', f'{score:.2f}', f'{fa_s_score:.2f}', f'{pu_s_s

    # Display scores per participant in a table
    print(tabulate(table_data, headers='firstrow', tablefmt='fancy_grid'))


    # Plot the scores of subscales
    subscale_titles = ['Overall Score', 'FA-S Score', 'PU-S Score', 'AE-S Score', 'RW-S Score']
    subscale_data = [ues_participant_scores, ues_subscale_scores['FA-S'], ues_subscale_scores['
    PlotScoresBoxPlot(subscale_titles, subscale_data, f'Boxplot of {description} Scores')

    # Calculate mean and standard deviation for each subscale
    print(f"----------- UES-SF mean and standard deviation for {description} ------------")
    print("\n\n")
```

```python
    print("------------ UES-SF Overall mean and standard deviation ------------")
    mean_overall, sd_overall = CalculateStats(ues_participant_scores, 'UES-SF Overall Engagemen
    
    print("------------ UES-SF FA-S mean and standard deviation ------------")
    mean_fa_s, sd_fa_s = CalculateStats(ues_subscale_scores['FA-S'], 'UES-SF FA-S Score')

    print("------------ UES-SF PU-S mean and standard deviation ------------")
    mean_pu_s, sd_pu_s = CalculateStats(ues_subscale_scores['PU-S'], 'UES-SF PU-S Score')

    print("------------ UES-SF AE-S mean and standard deviation ------------")
    mean_ae_s, sd_ae_s = CalculateStats(ues_subscale_scores['AE-S'], 'UES-SF AE-S Score')

    print("------------ UES-SF RW-S mean and standard deviation ------------")
    mean_rw_s, sd_rw_s = CalculateStats(ues_subscale_scores['RW-S'], 'UES-SF RW-S Score')

    means = [mean_overall, mean_fa_s, mean_pu_s, mean_ae_s, mean_rw_s]
    sds = [sd_overall, sd_fa_s, sd_pu_s, sd_ae_s, sd_rw_s]

    # Generate the table in LaTeX format
    # print(tabulate(table_data, headers='firstrow', tablefmt='latex'))

    return ues_participant_scores, ues_subscale_scores, means, sds
```

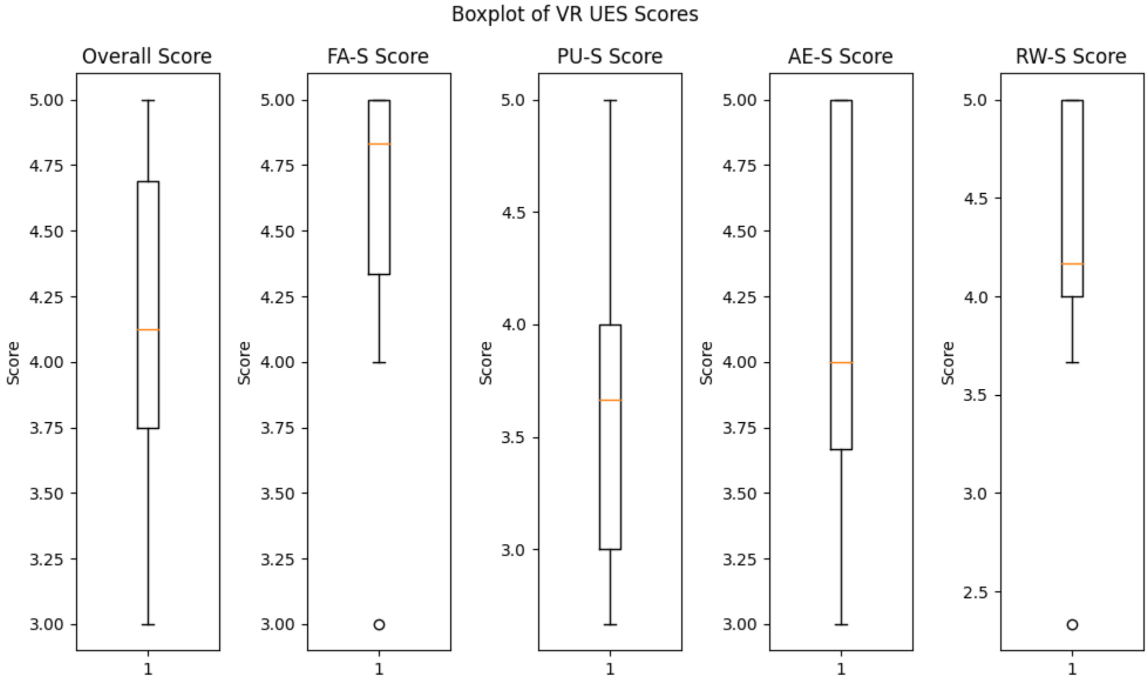
In [ ]:

```python
# VR UES items
ues_items_vr = {
    'FA-S.1': 'I lost myself in this experience.',
    'FA-S.2': 'The time I spent using this escpe room just slipped away.',
    'FA-S.3': 'I was absorbed in this experience.',
    'PU-S.1': 'I felt frustrated while using this escape room.',
    'PU-S.2': 'I found this escape room confusing to use.',
    'PU-S.3': 'Using this escape room was taxing.',
    'AE-S.1': 'The escape room was attractive.',
    'AE-S.2': 'The escape room was aesthetically applealing.',
    'AE-S.3': 'The escape room appealed to my senses.',
    'RW-S.1': 'Using the escape room was worthwile.',
    'RW-S.2': 'My experience was rewarding.',
    'RW-S.3': 'I felt interested in this experience.'
}

print("\n")

print("------------------------- VR UES Scores --------------------------------- ")
vr_ues_participant_scores, vr_ues_subscale_scores, vr_ues_means, vr_ues_sds = CalculateScoreUES

# Traditional UES items
ues_items = {
    'FA-S.1': 'I lost myself in this experience.',
    'FA-S.2': 'The time I spent using this online training just slipped away.',
    'FA-S.3': 'I was absorbed in this experience.',
    'PU-S.1': 'I felt frustrated while using this online training.',
    'PU-S.2': 'I found this online training  confusing to use.',
    'PU-S.3': 'Using this online training was taxing.',
    'AE-S.1': 'The online training was attractive.',
    'AE-S.2': 'The  online training was aesthetically applealing.',
    'AE-S.3': 'The  online training appealed to my senses.',
    'RW-S.1': 'Using the  online training was worthwile.',
    'RW-S.2': 'My experience was rewarding.',
    'RW-S.3': 'I felt interested in this experience.'
}

print("------------------------- Traditional UES Scores ------------------------- ")
traditional_ues_participant_scores, traditional_ues_subscale_scores, traditional_ues_means, tra
print("\n\n")

# Perform Mann-Whitney U test for overall UES scores
print("------------------------- Mann-Whitney U Test UES Overall Score ------------------------
PerformMannWhitneyUTest(vr_ues_participant_scores, traditional_ues_participant_scores)
print("\n\n")

# Perform Mann-Whitney U test for FA-S scores
print("------------------------- Mann-Whitney U Test UES FA-S Score ----------------------------
PerformMannWhitneyUTest(vr_ues_subscale_scores['FA-S'], traditional_ues_subscale_scores['FA-S']
print("\n\n")

# Perform Mann-Whitney U test for PU-S scores
```

```
print("------------------------- Mann-Whitney U Test UES PU-S Score ---------------------------
PerformMannWhitneyUTest(vr_ues_subscale_scores['PU-S'], traditional_ues_subscale_scores['PU-S']
print("\n\n")

# Perform Mann-Whitney U test for AE-S scores
print("------------------------- Mann-Whitney U Test UES AE-S Score ---------------------------
PerformMannWhitneyUTest(vr_ues_subscale_scores['AE-S'], traditional_ues_subscale_scores['AE-S']
print("\n\n")

# Perform Mann-Whitney U test for RW-S scores
print("------------------------- Mann-Whitney U Test UES RW-S Score ---------------------------
PerformMannWhitneyUTest(vr_ues_subscale_scores['RW-S'], traditional_ues_subscale_scores['RW-S']
print("\n\n")

# draw a table with the means and standard deviations of the scores for each subscale
table_data = [
    ['UES', 'VR Mean', 'VR SD', 'Traditional Mean', 'Traditional SD']
]

table_scale_titles = ['Overall', 'FA', 'PU', 'AE', 'RW']

# go through the overall and all subscales means and standard deviations and add them to the ta
for i, mean in enumerate(vr_ues_means):
    vr_sd = vr_ues_sds[i]
    traditional_mean = traditional_ues_means[i]
    traditional_sd = traditional_ues_sds[i]
    table_data.append([f'{table_scale_titles[i]}', f'{mean:.2f}', f'{vr_sd:.2f}', f'{traditiona

print("\n\n")
print("------------------------- VR vs Traditional - Means and Standard Deviations Comparison t
print(tabulate(table_data, headers='firstrow', tablefmt='fancy_grid'))

# generate latex table for the means and standard deviations
# print(tabulate(table_data, headers='firstrow', tablefmt='latex'))
```

------------------------- VR UES Scores ---------------------------------

Participant	Overall Engagement Score	FA-S Score	PU-S Score	AE-S Score	R W-S Score
Participant 1	5	5	5	5	5
Participant 2	4.83	5	4.33	5	5
Participant 3	4.75	5	4	5	5
Participant 4	3.75	4	3	4	4
Participant 5	4.25	5	3.67	3.67	4.67
Participant 6	4.25	4.67	4	4	4.33
Participant 7	4	5	3	4	4
Participant 8	5	5	5	5	5
Participant 9	3.67	5	2.67	3.33	3.67
Participant 10	3	3	3	3.67	2.33
Participant 11	3.92	4.33	3.33	4	4
Participant 12	4.08	4.33	3.67	4.67	3.67
Participant 13	4.17	4.67	3.33	4.67	4
Participant 14	3.67	4	3.67	3	4
Participant 15	4.67	5	3.67	5	5
Participant 16	3.75	4.33	2.67	3.33	4.67


Boxplot of VR UES Scores
Overall Score
FA-S Score
PU-S Score
AE-S Score
RW-S Score
Score
1

------------ UES-SF mean and standard deviation for VR UES ------------

------------ UES-SF Overall mean and standard deviation ------------

UES-SF Overall Engagement Score --> Mean: 4.17, Standard Deviation: 0.56

------------ UES-SF FA-S mean and standard deviation ------------

UES-SF FA-S Score --> Mean: 4.58, Standard Deviation: 0.56

------------ UES-SF PU-S mean and standard deviation ------------

UES-SF PU-S Score --> Mean: 3.62, Standard Deviation: 0.72

------------ UES-SF AE-S mean and standard deviation ------------

UES-SF AE-S Score --> Mean: 4.21, Standard Deviation: 0.70

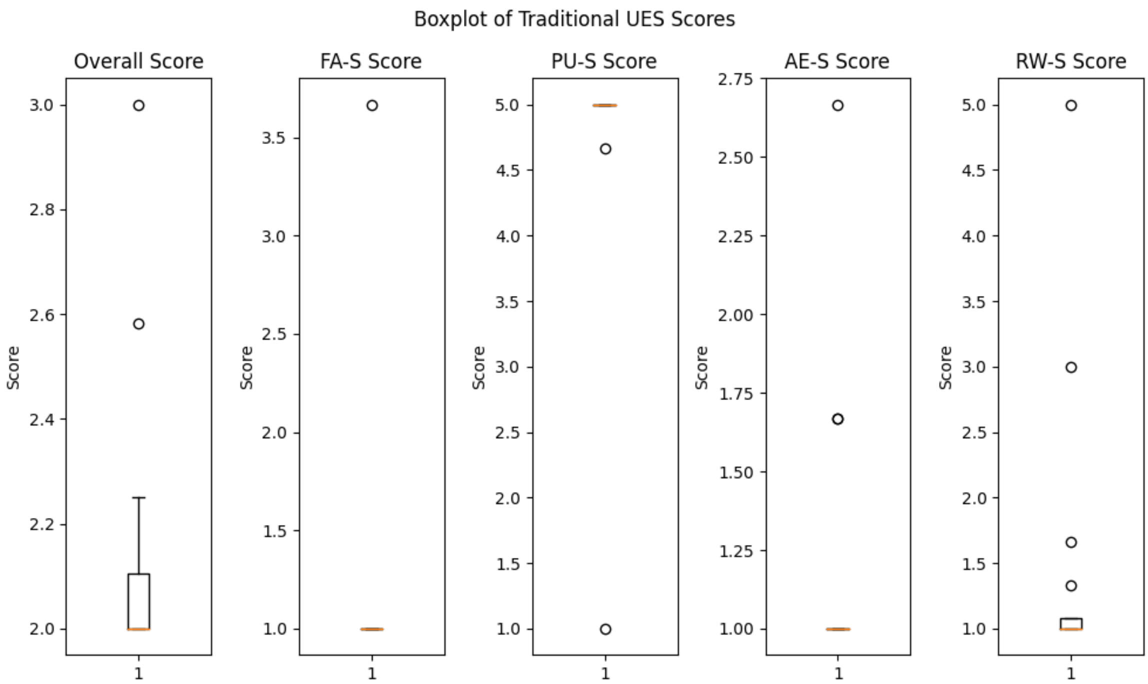
------------ UES-SF RW-S mean and standard deviation ------------

UES-SF RW-S Score --> Mean: 4.27, Standard Deviation: 0.72

------------------------- Traditional UES Scores -------------------------

Participant	Overall Engagement Score	FA-S Score	PU-S Score	AE-S Score	RW-S Score
Participant 1	2	1	5	1	1
Participant 2	2	1	5	1	1
Participant 3	2.08	1	5	1	1.33
Participant 4	2.17	1	5	1.67	1
Participant 5	2	1	5	1	1
Participant 6	2	1	5	1	1
Participant 7	2	1	5	1	1
Participant 8	2	1	5	1	1
Participant 9	2	1	5	1	1
Participant 10	3	1	5	1	5

Participant 11	2.25	1	4.67	1.67	1.67
Participant 12	2.58	3.67	1	2.67	3
Participant 13	2	1	5	1	1
Participant 14	2	1	5	1	1
Participant 15	2	1	5	1	1
Participant 16	2	1	5	1	1



```
----------- UES-SF mean and standard deviation for Traditional UES ------------


------------ UES-SF Overall mean and standard deviation ------------
UES-SF Overall Engagement Score --> Mean: 2.13, Standard Deviation: 0.28


------------ UES-SF FA-S mean and standard deviation ------------
UES-SF FA-S Score --> Mean: 1.17, Standard Deviation: 0.67


------------ UES-SF PU-S mean and standard deviation ------------
UES-SF PU-S Score --> Mean: 4.73, Standard Deviation: 1.00


------------ UES-SF AE-S mean and standard deviation ------------
UES-SF AE-S Score --> Mean: 1.19, Standard Deviation: 0.45


------------ UES-SF RW-S mean and standard deviation ------------
UES-SF RW-S Score --> Mean: 1.44, Standard Deviation: 1.08




------------------------- Mann-Whitney U Test UES Overall Score ----------------------------
Statistic: 255.5
p-value: 1.016851028552827e-06
Mann Whiney U Test --> Are the samples significantly different? -->  True


------------------------- Mann-Whitney U Test UES FA-S Score ----------------------------
Statistic: 255.0
p-value: 3.777654423790121e-07
Mann Whiney U Test --> Are the samples significantly different? -->  True


------------------------- Mann-Whitney U Test UES PU-S Score ----------------------------
Statistic: 32.0
p-value: 0.00011626825523632129
Mann Whiney U Test --> Are the samples significantly different? -->  True


------------------------- Mann-Whitney U Test UES AE-S Score ----------------------------
Statistic: 256.0
p-value: 6.012451936499951e-07
Mann Whiney U Test --> Are the samples significantly different? -->  True


------------------------- Mann-Whitney U Test UES RW-S Score ----------------------------
Statistic: 241.5
p-value: 1.0856098105063628e-05
Mann Whiney U Test --> Are the samples significantly different? -->  True




------------------------- VR vs Traditional - Means and Standard Deviations Comparison table ---
-------------------------
```

UES	VR Mean	VR SD	Traditional Mean	Traditional SD

Overall	4.17	0.56	2.13	0.28
FA	4.58	0.56	1.17	0.67
PU	3.62	0.72	4.73	1
AE	4.21	0.7	1.19	0.45
RW	4.27	0.72	1.44	1.08

Calculation of CAP Scores

The following function takes input data and a dictionary of CAP items as parameters. It calculates the total CAP score and subscale scores based on the CAP items, excluding the psychomotor subscale for the sake of this study's goals. The function reverses the scores for a specific item, calculates participant scores and subscale scores, and displays the results in a table format using the tabulate library. The function returns the total score, cognitive subscale scores, and affective subscale scores as output.

In [ ]:
```
def CalculateScoreCAP(data, cap_items, description):

    cap_columns = list(cap_items.values())

    cap_data = data[cap_columns].copy()

    # Reverse code items 2
    cap_data[cap_items[2]] = 6 - cap_data[cap_items[2]]

    # Calculate total CAP score
    cap_data['Total Score'] = cap_data.sum(axis=1)

    # Calculate subscale scores
    cap_data['Cognitive Subscale'] = cap_data[[cap_items[1], cap_items[2], cap_items[5]]].sum(a
    cap_data['Affective Subscale'] = cap_data[[cap_items[4], cap_items[6], cap_items[9]]].sum(a

    # Prepare the data for the table
    table_data = [
        ['Participant', 'Total Score', 'Cognitive Subscale', 'Affective Subscale']
    ]

    for i, score in enumerate(cap_data['Total Score']):
        cognitive_score = cap_data['Cognitive Subscale'].iloc[i]
        affective_score = cap_data['Affective Subscale'].iloc[i]
        table_data.append([f'Participant {i+1}', score, cognitive_score, affective_score])

    # Display scores per participant in a table
    print(tabulate(table_data, headers='firstrow', tablefmt='fancy_grid'))

    maxTotalScore = 54 - (6 * 3)  # we removed 3 items from the scale: items number 3,7 and 8 (
    print(f'Maximum Total Score: {maxTotalScore}')

    # Generate the table in LaTeX format
    # print(tabulate(table_data, headers='firstrow', tablefmt='latex'))

    # Plot the scores of subscales
    subscale_titles = ['Total Score', 'Cognitive Subscale', 'Affective Subscale']
    subscale_data = [cap_data['Total Score'], cap_data['Cognitive Subscale'], cap_data['Affecti
    PlotScoresBoxPlot(subscale_titles, subscale_data, f'Boxplot of {description} Scores')

    # Calculate mean and standard deviation for each subscale
    print(f"----------- CAP mean and standard deviation for {description} ------------")
    print("\n\n")
    print("------------ CAP Total Score mean and standard deviation ------------")
    mean_total, sd_total = CalculateStats(cap_data['Total Score'], 'CAP Total Score')

    print("------------ CAP Cognitive Subscale mean and standard deviation ------------")
    mean_cognitive, sd_cognitive = CalculateStats(cap_data['Cognitive Subscale'], 'CAP Cognitiv

    print("------------ CAP Affective Subscale mean and standard deviation ------------")
    mean_affective, sd_affective = CalculateStats(cap_data['Affective Subscale'], 'CAP Affectiv

    means = [mean_total, mean_cognitive, mean_affective]
```

```
    sds = [sd_total, sd_cognitive, sd_affective]

    return cap_data['Total Score'], cap_data['Cognitive Subscale'], cap_data['Affective Subscal
```

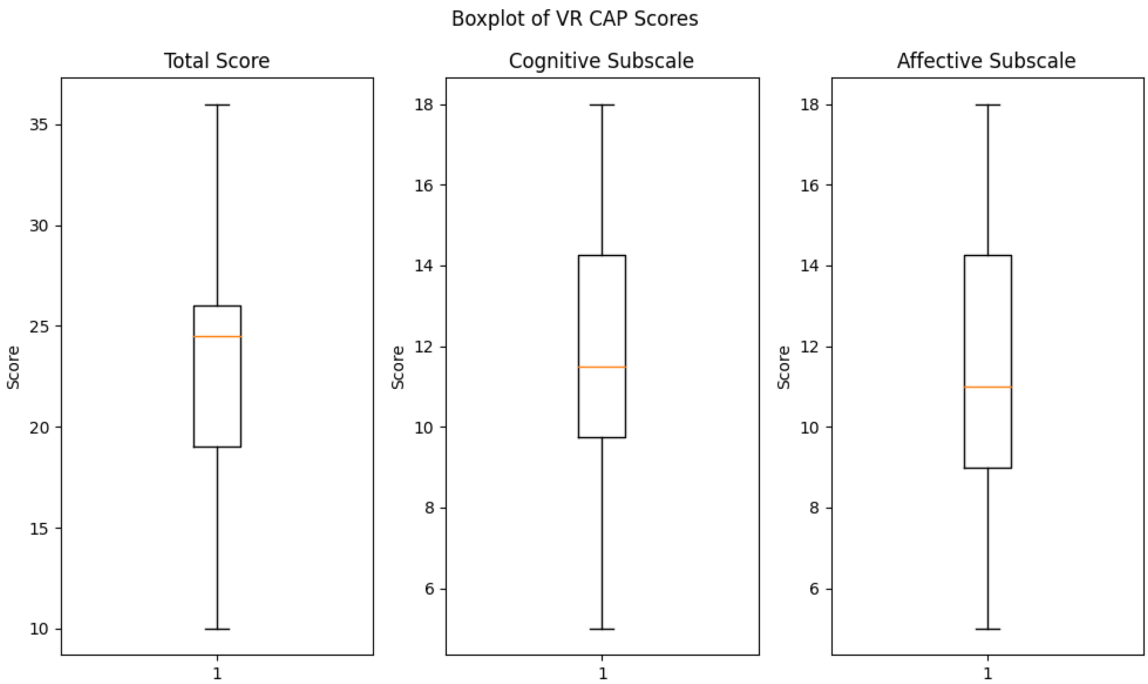
In [ ]:

```
cap_items_vr = {
    1: 'I can organize the training material into a logical structure.',
    2: 'I cannot produce a training study guide for future students.',
    4: 'I have changed my attitudes about the training subject matter as a result of this train
    5: 'I can intelligently critique the material used in this training.',
    6: 'I feel more self-reliant as the result of the content learned in this training.',
    9: 'I feel that I am a more sophisticated thinker as a result of this training.'
}

print("------------------------- VR CAP Scores --------------------------------- ")
vr_cap_total_score, vr_cap_cognitive_scores, vr_cap_affective_scores, vr_cap_means, vr_cap_sds

print("\n")

cap_items_traditional = {
    1: 'I can organize the training material into a logical structure.',
    2: 'I cannot produce a training study guide for future students.',
    4: 'I have changed my attitudes about the training subject matter as a result of this train
    5: 'I can intelligently critique the material used in this training.',
    6: 'I feel more self-reliant as the result of the content learned in this training.',
    9: 'I feel that I am a more sophisticated thinker as a result of this training.'
}

print("------------------------- Traditional CAP Scores ------------------------- ")
traditional_cap_total_score, traditional_cap_cognitive_scores, traditional_cap_affective_scores

print("\n\n")

# Perform Mann-Whitney U test for CAP total scores
print("------------------------- Mann-Whitney U Test CAP Total Score --------------------------
PerformMannWhitneyUTest(vr_cap_total_score, traditional_cap_total_score)
print("\n\n")

# Perform Mann-Whitney U test for CAP cognitive scores
print("------------------------- Mann-Whitney U Test CAP Cognitive Score ----------------------
PerformMannWhitneyUTest(vr_cap_cognitive_scores, traditional_cap_cognitive_scores)
print("\n\n")

# Perform Mann-Whitney U test for CAP affective scores
print("------------------------- Mann-Whitney U Test CAP Affective Score ----------------------
PerformMannWhitneyUTest(vr_cap_affective_scores, traditional_cap_affective_scores)
print("\n\n")


# draw a table with the means and standard deviations of the scores for each subscale
table_data = [
    ['CAP', 'VR Mean', 'VR SD', 'Traditional Mean', 'Traditional SD']
]

table_scale_titles = ['Total', 'Cognitive', 'Affective']

# go through the overall and all subscales means and standard deviations and add them to the ta
for i, mean in enumerate(vr_cap_means):
    vr_sd = vr_cap_sds[i]
    traditional_mean = traditional_cap_means[i]
    traditional_sd = traditional_cap_sds[i]
    table_data.append([f'{table_scale_titles[i]}', f'{mean:.2f}', f'{vr_sd:.2f}', f'{traditiona

print("\n\n")
print("------------------------- VR vs Traditional - Means and Standard Deviations Comparison t
print(tabulate(table_data, headers='firstrow', tablefmt='fancy_grid'))

# generate latex table for the means and standard deviations
# print(tabulate(table_data, headers='firstrow', tablefmt='latex'))
```

------------------------ VR CAP Scores --------------------------------

Participant	Total Score	Cognitive Subscale	Affective Subscale
Participant 1	36	18	18
Participant 2	25	13	12
Participant 3	36	18	18
Participant 4	25	14	11
Participant 5	24	10	14
Participant 6	26	11	15
Participant 7	18	12	6
Participant 8	28	15	13
Participant 9	10	5	5
Participant 10	19	9	10
Participant 11	18	9	9
Participant 12	19	11	8
Participant 13	25	10	15
Participant 14	21	12	9
Participant 15	26	15	11
Participant 16	19	9	10

Maximum Total Score: 36



----------- CAP mean and standard deviation for VR CAP ------------

------------ CAP Total Score mean and standard deviation ------------
CAP Total Score --> Mean: 23.44, Standard Deviation: 6.64

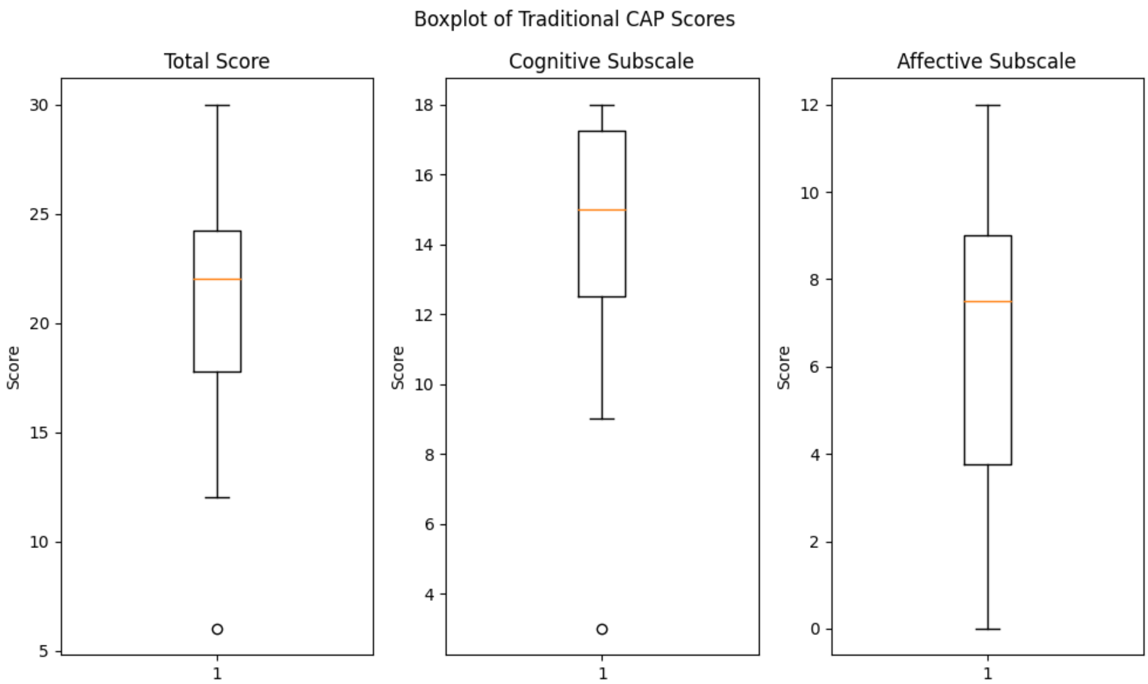
------------ CAP Cognitive Subscale mean and standard deviation ------------
CAP Cognitive Subscale --> Mean: 11.94, Standard Deviation: 3.47

------------ CAP Affective Subscale mean and standard deviation ------------
CAP Affective Subscale --> Mean: 11.50, Standard Deviation: 3.83

------------------------- Traditional CAP Scores -------------------------

Participant	Total Score	Cognitive Subscale	Affective Subscale
Participant 1	25	15	10
Participant 2	15	15	0
Participant 3	17	14	3
Participant 4	12	10	2
Participant 5	22	16	6
Participant 6	24	18	6
Participant 7	22	18	4
Participant 8	24	18	6
Participant 9	20	11	9
Participant 10	30	18	12
Participant 11	18	9	9
Participant 12	6	3	3
Participant 13	24	15	9
Participant 14	22	13	9
Participant 15	28	16	12
Participant 16	26	17	9

Maximum Total Score: 36


Boxplot of Traditional CAP Scores
Total Score
Cognitive Subscale
Affective Subscale
Score
30
25
20
15
10
5
18
16
14
12
10
8
6
4
12
10
8
6
4
2
0
1

```
----------- CAP mean and standard deviation for Traditional CAP ------------


------------ CAP Total Score mean and standard deviation ------------
CAP Total Score --> Mean: 20.94, Standard Deviation: 6.16


------------ CAP Cognitive Subscale mean and standard deviation ------------
CAP Cognitive Subscale --> Mean: 14.12, Standard Deviation: 4.13


------------ CAP Affective Subscale mean and standard deviation ------------
CAP Affective Subscale --> Mean: 6.81, Standard Deviation: 3.62




------------------------- Mann-Whitney U Test CAP Total Score ----------------------------
Statistic: 152.5
p-value: 0.3640364599015392
Mann Whiney U Test --> Are the samples significantly different? -->  False


------------------------- Mann-Whitney U Test CAP Cognitive Score ----------------------------
Statistic: 78.5
p-value: 0.06289569711207751
Mann Whiney U Test --> Are the samples significantly different? -->  False


------------------------- Mann-Whitney U Test CAP Affective Score ----------------------------
Statistic: 206.5
p-value: 0.0030666469630165347
Mann Whiney U Test --> Are the samples significantly different? -->  True




------------------------- VR vs Traditional - Means and Standard Deviations Comparison table ---
-------------------------
```

CAP	VR Mean	VR SD	Traditional Mean	Traditional SD
Total	23.44	6.64	20.94	6.16
Cognitive	11.94	3.47	14.12	4.13
Affective	11.5	3.83	6.81	3.62

Calculation of SIMS Scores

The following function takes input data and a dictionary of SIMS items as parameters. It calculates the subscale means for intrinsic motivation, identified regulation, external regulation, and amotivation based on the SIMS items. The function prepares the data for a table display and shows the scores per participant. It returns the calculated subscale means for intrinsic motivation, identified regulation, external regulation, and amotivation as output.

In [ ]:
```
def CalculateScoreSIMS(data, sims_items, description):
    sims_columns = list(sims_items.values())
    sims_data = data[sims_columns]

    # Calculate subscale means
    intrinsic_items = [1, 5, 9, 13]
```

```python
    identified_items = [2, 6, 10, 14]
    external_items = [3, 7, 11, 15]
    amotivation_items = [4, 8, 12, 16]

    intrinsic_mean = sims_data[[sims_items[item] for item in intrinsic_items]].mean(axis=1)
    identified_mean = sims_data[[sims_items[item] for item in identified_items]].mean(axis=1)
    external_mean = sims_data[[sims_items[item] for item in external_items]].mean(axis=1)
    amotivation_mean = sims_data[[sims_items[item] for item in amotivation_items]].mean(axis=1)

    # Prepare the data for the table
    table_data = [
        ['Participant', 'Intrinsic Motivation', 'Identified Regulation', 'External Regulation',
    ]

    for i, score in enumerate(sims_data):
        intrinsic = intrinsic_mean.iloc[i]
        identified = identified_mean.iloc[i]
        external = external_mean.iloc[i]
        amotivation = amotivation_mean.iloc[i]
        table_data.append([f'Participant {i+1}', intrinsic, identified, external, amotivation])

    # Display scores per participant in a table
    print(tabulate(table_data, headers='firstrow', tablefmt='fancy_grid'))

    # Generate the table in LaTeX format
    # print(tabulate(table_data, headers='firstrow', tablefmt='latex'))

    # Plot the scores of subscales
    subscale_titles = ['Intrinsic Motivation', 'Identified Regulation', 'External Regulation',
    subscale_data = [intrinsic_mean, identified_mean, external_mean, amotivation_mean]
    PlotScoresBoxPlot(subscale_titles, subscale_data, f'Boxplot of {description} Scores')

    # Calculate mean and standard deviation for each subscale
    print(f"----------- SIMS mean and standard deviation for {description} ------------")
    print("\n\n")
    print("------------ SIMS Intrinsic Motivation mean and standard deviation ------------")
    mean_intrinsic, sd_intrinsic = CalculateStats(intrinsic_mean, 'SIMS Intrinsic Motivation')

    print("------------ SIMS Identified Regulation mean and standard deviation ------------")
    mean_identified, sd_identified = CalculateStats(identified_mean, 'SIMS Identified Regulatio

    print("------------ SIMS External Regulation mean and standard deviation ------------")
    mean_external, sd_external = CalculateStats(external_mean, 'SIMS External Regulation')

    print("------------ SIMS Amotivation mean and standard deviation ------------")
    mean_amotivation, sd_amotivation = CalculateStats(amotivation_mean, 'SIMS Amotivation')

    means = [mean_intrinsic, mean_identified, mean_external, mean_amotivation]
    sds = [sd_intrinsic, sd_identified, sd_external, sd_amotivation]

    return intrinsic_mean, identified_mean, external_mean, amotivation_mean, means, sds
```

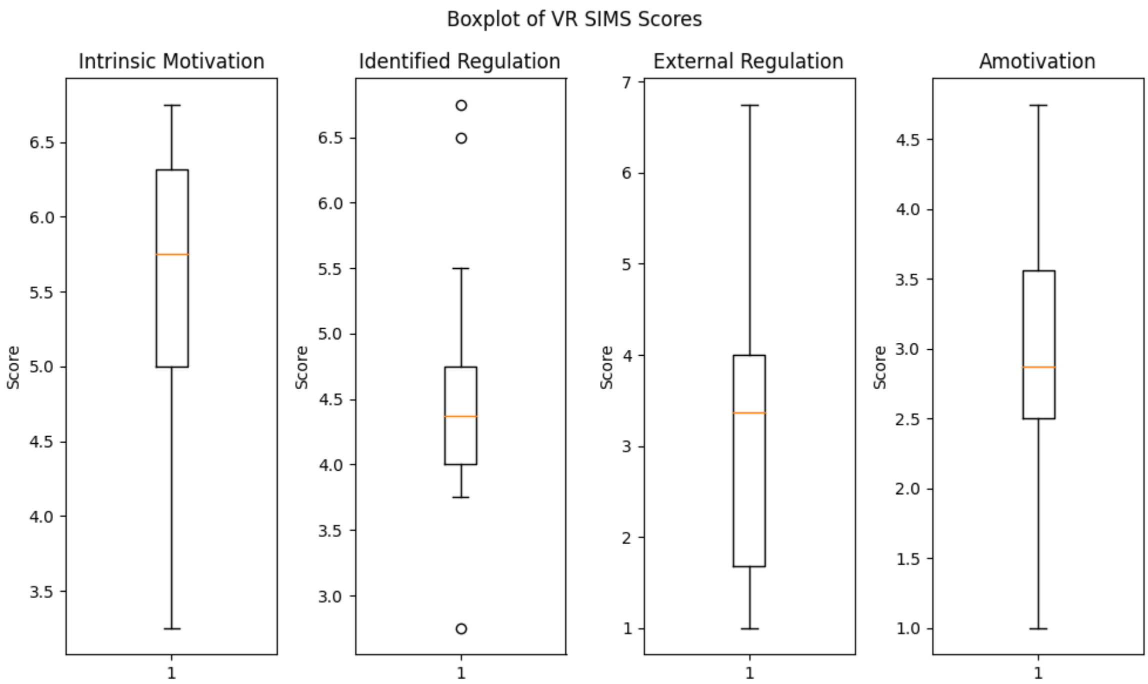
In [ ]:

```python
sims_items_vr = {
    1: 'Because I think that this activity is interesting.',
    2: 'Because I am doing it for my own good.',
    3: 'Because I am supposed to do it.',
    4: 'There may be good reasons to do this activity, but personally I don’t see any.',
    5: 'Because I think that this activity is pleasant.',
    6: 'Because I think that this activity is good for me.',
    7: 'Because it is something that I have to do.',
    8: 'I do this activity but I am not sure if it is worth it.',
    9: 'Because this activity is fun.',
    10: 'By personal decision.',
    11: 'Because I don’t have any choice.',
    12: 'I don’t know; I don’t see what this activity brings me.',
    13: 'Because I feel good when doing this activity.',
    14: 'Because I believe that this activity is important for me.',
    15: 'Because I feel that I have to do it.',
    16: 'I do this activity, but I am not sure it is a good thing to pursue it.'
}

print("------------------------- VR SIMS Scores ---------------------------------- ")
vr_sims_intrinsic_scores, vr_sims_identified_scores, vr_sims_external_scores, vr_sims_amotivati

print("\n")
```

```python
sims_items_traditional = {
    1: 'Because I think that this activity is interesting.',
    2: 'Because I am doing it for my own good.',
    3: 'Because I am supposed to do it.',
    4: 'There may be good reasons to do this activity, but personally I don’t see any.',
    5: 'Because I think that this activity is pleasant.',
    6: 'Because I think that this activity is good for me.',
    7: 'Because it is something that I have to do.',
    8: 'I do this activity but I am not sure if it is worth it.',
    9: 'Because this activity is fun.',
    10: 'By personal decision.',
    11: 'Because I don’t have any choice.',
    12: 'I don’t know; I don’t see what this activity brings me.',
    13: 'Because I feel good when doing this activity.',
    14: 'Because I believe that this activity is important for me.',
    15: 'Because I feel that I have to do it.',
    16: 'I do this activity, but I am not sure it is a good thing to pursue it.'
}

print("------------------------- Traditional SIMS Scores ------------------------- ")
traditional_sims_intrinsic_scores, traditional_sims_identified_scores, traditional_sims_externa

print("\n\n")

# Perform Mann-Whitney U test for SIMS intrinsic scores
print("------------------------- Mann-Whitney U Test SIMS Intrinsic Score ----------------------
PerformMannWhitneyUTest(vr_sims_intrinsic_scores, traditional_sims_intrinsic_scores)
print("\n\n")

# Perform Mann-Whitney U test for SIMS identified scores
print("------------------------- Mann-Whitney U Test SIMS Identified Score ---------------------
PerformMannWhitneyUTest(vr_sims_identified_scores, traditional_sims_identified_scores)
print("\n\n")

# Perform Mann-Whitney U test for SIMS external scores
print("------------------------- Mann-Whitney U Test SIMS External Score -----------------------
PerformMannWhitneyUTest(vr_sims_external_scores, traditional_sims_external_scores)
print("\n\n")

# Perform Mann-Whitney U test for SIMS amotivation scores
print("------------------------- Mann-Whitney U Test SIMS Amotivation Score --------------------
PerformMannWhitneyUTest(vr_sims_amotivation_scores, traditional_sims_amotivation_scores)
print("\n\n")


# draw a table with the means and standard deviations of the scores for each subscale
table_data = [
    ['SIMS', 'VR Mean', 'VR SD', 'Traditional Mean', 'Traditional SD']
]

table_scale_titles = ['Intrinsic', 'Identified', 'External', 'Amotivation']

# go through the overall and all subscales means and standard deviations and add them to the ta
for i, mean in enumerate(vr_sims_means):
    vr_sd = vr_sims_sds[i]
    traditional_mean = traditional_sims_means[i]
    traditional_sd = traditional_sims_sds[i]
    table_data.append([f'{table_scale_titles[i]}', f'{mean:.2f}', f'{vr_sd:.2f}', f'{traditiona

print("\n\n")
print("------------------------- VR vs Traditional - Means and Standard Deviations Comparison t
print(tabulate(table_data, headers='firstrow', tablefmt='fancy_grid'))

# generate latex table for the means and standard deviations
# print(tabulate(table_data, headers='firstrow', tablefmt='latex'))
```

------------------------ VR SIMS Scores --------------------------------

Participant	Intrinsic Motivation	Identified Regulation	External Regulation	Amotivation
Participant 1	6	4	4	3.5
Participant 2	5	4.5	3.5	3
Participant 3	6.25	5.5	3	3
Participant 4	5.25	4	4.5	3.75
Participant 5	6.5	4.5	3.25	2.75
Participant 6	6.75	5.5	1.75	2.5
Participant 7	5.75	4	1	3.25
Participant 8	6.75	6.75	1.5	1
Participant 9	5.75	3.75	6	4.75
Participant 10	5	4.25	4	4.25
Participant 11	5	4.25	3.5	2.5
Participant 12	5.25	4.5	4	2.75
Participant 13	3.25	2.75	6.75	3.75
Participant 14	5.75	4.5	1.5	2.25
Participant 15	6.75	6.5	1	1
Participant 16	5	3.75	2.75	2.5



Boxplot of VR SIMS Scores
Intrinsic Motivation
Identified Regulation
External Regulation
Amotivation
Score
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
7
6
5
4
3
2
1
2.5
2.0
1.5
1.0

----------- SIMS mean and standard deviation for VR SIMS ------------

------------ SIMS Intrinsic Motivation mean and standard deviation ------------
SIMS Intrinsic Motivation --> Mean: 5.62, Standard Deviation: 0.92

------------ SIMS Identified Regulation mean and standard deviation ------------
SIMS Identified Regulation --> Mean: 4.56, Standard Deviation: 1.03

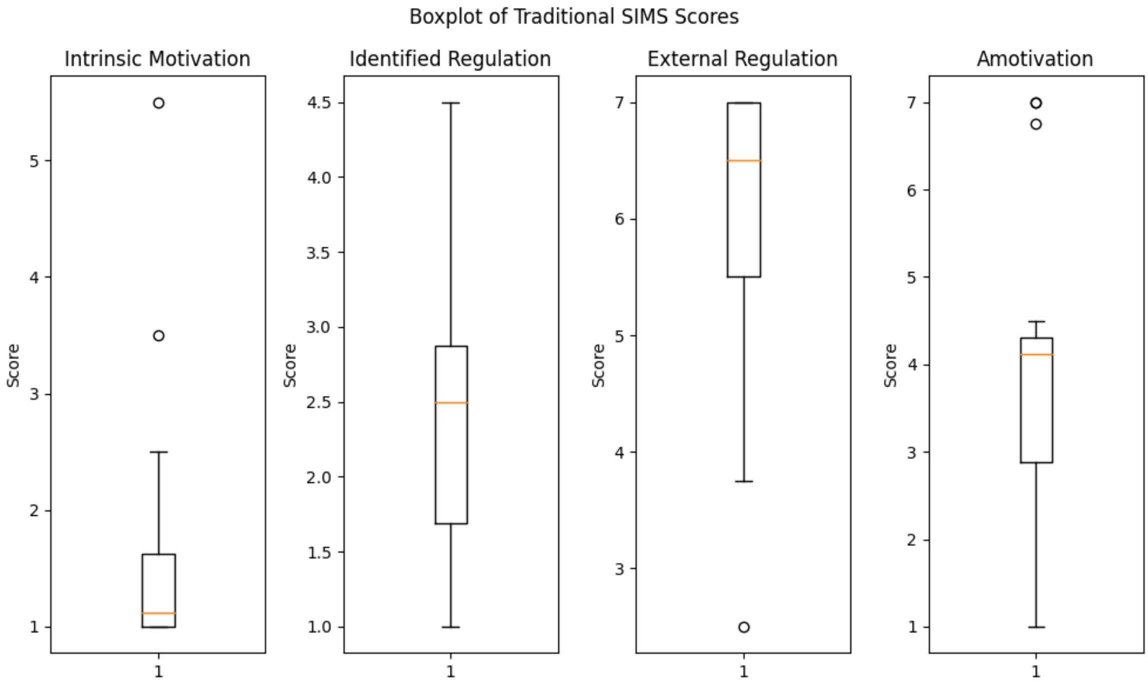
------------ SIMS External Regulation mean and standard deviation ------------
SIMS External Regulation --> Mean: 3.25, Standard Deviation: 1.67

------------ SIMS Amotivation mean and standard deviation ------------
SIMS Amotivation --> Mean: 2.91, Standard Deviation: 1.01

------------------------- Traditional SIMS Scores -------------------------

Participant	Intrinsic Motivation	Identified Regulation	External Regulation	Amotivation
Participant 1	1.5	2.75	5.5	4.25
Participant 2	1	1	7	1
Participant 3	1.25	1.5	7	3
Participant 4	1.25	1.75	6.75	4.25
Participant 5	1	1.75	6.25	6.75
Participant 6	1	1	7	7
Participant 7	1.5	2.5	7	3.25
Participant 8	1	1	7	7
Participant 9	1	3.75	7	2.5
Participant 10	5.5	2.5	2.5	1
Participant 11	2.5	4.5	3.75	4.25

Participant 12	3.5	3.25	5	4
Participant 13	1	1.75	7	4.5
Participant 14	1	2.5	6.25	4
Participant 15	1	2.5	6	4.25
Participant 16	2	4.5	5.5	2.5



```
----------- SIMS mean and standard deviation for Traditional SIMS ------------


------------ SIMS Intrinsic Motivation mean and standard deviation ------------
SIMS Intrinsic Motivation --> Mean: 1.69, Standard Deviation: 1.23


------------ SIMS Identified Regulation mean and standard deviation ------------
SIMS Identified Regulation --> Mean: 2.41, Standard Deviation: 1.14


------------ SIMS External Regulation mean and standard deviation ------------
SIMS External Regulation --> Mean: 6.03, Standard Deviation: 1.33


------------ SIMS Amotivation mean and standard deviation ------------
SIMS Amotivation --> Mean: 3.97, Standard Deviation: 1.83





------------------------- Mann-Whitney U Test SIMS Intrinsic Score ----------------------------
Statistic: 248.0
p-value: 5.420982007379307e-06
Mann Whiney U Test --> Are the samples significantly different? -->  True


------------------------- Mann-Whitney U Test SIMS Identified Score ----------------------------
Statistic: 232.5
p-value: 8.052208810713059e-05
Mann Whiney U Test --> Are the samples significantly different? -->  True


------------------------- Mann-Whitney U Test SIMS External Score ----------------------------
Statistic: 27.0
p-value: 0.00013828850877508158
Mann Whiney U Test --> Are the samples significantly different? -->  True


------------------------- Mann-Whitney U Test SIMS Amotivation Score ---------------------------
-
Statistic: 76.5
p-value: 0.05328393268545082
Mann Whiney U Test --> Are the samples significantly different? -->  False




------------------------- VR vs Traditional - Means and Standard Deviations Comparison table ---
-------------------------
```

SIMS	VR Mean	VR SD	Traditional Mean	Traditional SD
Intrinsic	5.62	0.92	1.69	1.23
Identified	4.56	1.03	2.41	1.14
External	3.25	1.67	6.03	1.33
Amotivation	2.91	1.01	3.97	1.83

In [ ]:

```
# get the additional questions data from the vr data CSV file
# they  are:
```

```
# I liked this escape room better than a traditional self-study online courses.
# I learned more with this escape room than I would have with a traditional self-study online c
# I felt like I was part of the team.
# I felt like I was contributing to the team.
# My actions had an impact on the team.
# The actions of other team members affected me.
# I felt left out.

additional_questions_vr_items = [
    'I liked this escape room better than a traditional self-study online courses.',
    'I learned more with this escape room than I would have with a traditional self-study onlin
    'I felt like I was part of the team.',
    'I felt like I was contributing to the team.',
    'My actions had an impact on the team.',
    'The actions of other team members affected me.',
    'I felt left out.'
]

# calculate the mean and standard deviation for each additional question
additional_questions_vr_means = []
additional_questions_vr_sds = []

for item in additional_questions_vr_items:
    mean, sd = CalculateStats(data_vr[item], item)
    additional_questions_vr_means.append(mean)
    additional_questions_vr_sds.append(sd)


# make a table with the means and standard deviations of the additional questions
table_data = [
    ['Additional Questions', 'Mean', 'SD']
]

# go through the means and standard deviations and add them to the table
for i, mean in enumerate(additional_questions_vr_means):
    sd = additional_questions_vr_sds[i]
    table_data.append([f'{additional_questions_vr_items[i]}', f'{mean:.2f}', f'{sd:.2f}'])

print("\n\n")

print("------------------------- VR Additional Questions - Means and Standard Deviations ------
print(tabulate(table_data, headers='firstrow', tablefmt='fancy_grid'))

# generate latex table for the means and standard deviations
# print(tabulate(table_data, headers='firstrow', tablefmt='latex'))
```

```
I liked this escape room better than a traditional self-study online courses. --> Mean: 4.56, Standard Deviation: 0.73


I learned more with this escape room than I would have with a traditional self-study online courses. --> Mean: 3.88, Standard Deviation: 1.15


I felt like I was part of the team. --> Mean: 4.69, Standard Deviation: 0.48


I felt like I was contributing to the team. --> Mean: 4.50, Standard Deviation: 0.89


My actions had an impact on the team. --> Mean: 4.50, Standard Deviation: 0.89


The actions of other team members affected me. --> Mean: 4.62, Standard Deviation: 0.62


I felt left out. --> Mean: 1.75, Standard Deviation: 0.93
```

```
------------------------- VR Additional Questions - Means and Standard Deviations ----------------------------
```

Additional Questions	Mean	SD
I liked this escape room better than a traditional self-study online courses.	4.56	0.73
I learned more with this escape room than I would have with a traditional self-study online courses.	3.88	1.15
I felt like I was part of the team.	4.69	0.48
I felt like I was contributing to the team.	4.5	0.89
My actions had an impact on the team.	4.5	0.89
The actions of other team members affected me.	4.62	0.62
I felt left out.	1.75	0.93

In [ ]: