MASTER THESIS

Designing a personalized feedback strategy for the Space Fortress game

By:

Iris J. Kroos

University of Twente

Educational Science & Technology

Faculty of Behavioral, Management and Social Sciences (BMS)

June 2021

Supervisors:

First: prof. dr. ir. B.P. Veldkamp

Second: dr. P. Papadopoulos

External: D. Thijssen MSc

External organization:

Netherlands Aerospace Centre (NLR)

Dedicated to innovation in aerospace

UNIVERSITY OF TWENTE.

Table of contents

Acknowledgment

Writing a master thesis during the COVID-19 pandemic was not an easy task to complete. Gladly I received the guidance of multiple people to successfully achieve this challenge. First, I would like to thank my supervisors prof. dr. Bernard Veldkamp and dr. Pantelis Papadopoulos for their clear guidance and feedback during this study. They sharpened my thinking and taught me how to bring this research to a higher level.

Second, I would like to thank my colleagues from my internship at NLR, for their support and help during this study. I enjoyed working together. I would particularly thank my supervisor at NLR, Dirk Thijssen. Dirk, thank you for your guidance, patience, feedback, and trust during this study. You helped me to become a better researcher.

Finally, I would like to thank my family and friends for their support during this study, they provided wise counsel and distractions when needed.

Abstract

The Royal Netherlands Aerospace Centre (NLR) made a revised version of the Space Fortress game, which functions as a training instrument to enhance gaming performance and as a measurement instrument for skill decay research. Participants learn to play the game and improve their skills. The skill decay is determined by performance measurement over different retention periods. However, skill decay is also associated with the absence of feedback, which is not included in the current version of the game. Therefore, NLR wants to apply personalized feedback in the training part of the game. In the ideal state of affairs, each participant of the Space Fortress game receives personalized feedback that improves their performance.

This research aimed to investigate which learner characteristics were connected with participants' performances in the Space Fortress game. Based upon this analysis a personalized feedback strategy was suggested. To investigate which characteristics of a personalized feedback strategy can be used to develop personalized feedback for the Space Fortress game, the theory of the feedback models of Mason and Bruning (2001) and Narciss and Huth (2004) were explored. The research was conducted using the learner characteristic, experience, and performance data of 10 participants who played the revised Space Fortress game training of NLR. The participants responded to multiple learner measurements questionnaires and were observed by the researcher. Responses were analyzed with the use of thematic analysis, Pearson correlation coefficient, and exploratory analysis. The results of the thematic analysis showed the commonly made errors, incorrect strategies, and potential learning problems. Pearson correlation coefficient showed that overall positive but weak correlations were found between motivation and performance scores and self-efficacy and performance scores. Positive strong correlations were found between gaming frequency and participants' performances. The exploratory analysis showed that based on the learner characteristics and performances three categories of learners can be created.

The results suggest that the different categories of participants had different learner characteristics and performance outcomes, and encounter different errors, incorrect strategies, and learning problems regarding playing the game. On this basis, the concept of participants' learner characteristics, performances, and experiences were taken into account to determine the function, presentation, and content of the feedback.

1. Introduction

To successfully perform in safety-critical professions, professionals need to maintain their performance and proficiency in complex skills (Vlasblom, Pennings, van der Pal, & Oprins, 2020). However, decreased performance frequently occurs since individuals often experience difficulties recalling certain skills, due to infrequent use (Kim, Ritter, & Koubek, 2013). The so-called 'skill decay' refers to the decay or loss of required or trained skills (or knowledge) after periods of nonuse (Arthur Jr, Bennett Jr, Stanush, & McNelly, 1998; Kluge & Frank, 2014).

According to Kim et al. (2013) skill decay is in particular significant in professions where individuals must successfully perform important skills that they rarely practice. In first responder aviation, medical and military contexts skill decay can have disastrous consequences, people can die or get injured as a result of task performance of these professionals (Vlasblom et al., 2020). Therefore, the proficiency of skills is crucial in safety-critical professions, even when the skill is not used for a long period of time (Kluge & Frank, 2014; Vlasblom et al., 2020). It is important to frequently recover acquired knowledge and developed skills to avoid decay. However, current training curricula do not differentiate between individuals. This means that regarding how much effort it takes, every individual performs the same content and hours of training (Lieffijn, 2020). No extensive research has been conducted about the input for professional refresher training, one reason for this limited scientific evidence is that the current literature focuses on retention of elementary skills or knowledge, training for professionals requires more insights into the retention of complex skills (Vlasblom et al., 2020).

For this reason, the Royal Netherlands Aerospace Centre (NLR) started a research project on skill retention or (alternatively skill decay) of complex skills for highly skilled professionals. The NLR skill decay research aims to create a personalized model of skill retention, that can predict the optimal refresher training moment for a specific person (professional). To built this model, performance data of a complex learning task is collected via an adaptive instructional system (AIS) (Van der Pal & Toubman, 2020). This AIS is built around the Space Fortress game and called the Space Fortress Adaptive Instructional System (SF-AIS). The SF-AIS is used as a training and measurement instrument where participants learn to play the Space Fortress game, which is a complex learning task. The Space Fortress game aims to improve the particular set of skills necessary to play the game.

To measure skill decay, participants' performance is measured over different retention periods. However, knowledge or skill loss has not only been associated with longer retention periods. Arthur Jr et al. (1998) stated that knowledge or skill loss has also been associated with absent or inadequate feedback.

When retention periods do not include deliberate practice with refresher learning opportunities or feedback, skills may no longer be functional (Weaver, Newman-Toker, & Rosen, 2012). However, the current training element of the SF-AIS is lacking a feedback mechanism, the only feedback provided relates to the knowledge of results of the individual games, only end scores are shown. This appears to be inadequate. To improve participant's performances and prevent possible decay or loss of skills due to the lack of feedback, NLR wants to provide the participants of the SF-AIS training with personalized feedback that guides their learning process.

To accomplish this, the training element of the SF-AIS needs to contain feedback. Therefore, a system has to be developed to create feedback for the training element of the SF-AIS. In support of the NLR skill decay research, this study aimed to investigate which characteristics of a personalized feedback strategy could increase participants' performance in the Space Fortress game. Several studies created frameworks and guidelines for feedback strategies (Narciss, 2013; Narciss & Huth, 2004; Shute, 2008). These frameworks identify important characteristics of a personalized feedback strategy aimed at designing adaptive formative and summative feedback, upon which this study focused. Suggestions for the design of personalized feedback can be made by identifying the individual learner characteristics, the learning and performance indicators, and homogeneous groups of the participants.

2. Theoretical framework

2.1 Feedback in educational games

Multiple studies stated that playing a game in a game-based learning environment does not necessarily lead to learning, instead there is a need for clear guidance and instruction to inform what information is needed for learning processes to take place (Mayer & Johnson, 2010; Serge, Priest, Durlach, & Johnson, 2013). To ensure learning, feedback is widely accepted to help shape the cognition, perception, or action of the learner (Killingsworth, Clark, & Adams, 2015; Serge et al., 2013). According to Charles, Charles, McNeill, Bustard, and Black (2011) feedback is fundamental to the process of gameplaying. First of all, feedback can function as an advanced organizer by providing learning guidance, suggestions about meaningful organization of the content to be learned, and by stating objectives (Cameron & Dwyer, 2005; Serge et al., 2013). Second, feedback leads to better learning motivation and can positively enhance the willingness for a participant to continue learning (Burgers, Eden, van Engelenburg, & Buningh, 2015; Corbalan, Kester, & van Merriënboer, 2009; Erhel & Jamet, 2013). One reason for the motivational aspect of feedback is that feedback promotes the relevance of the learning material, it enables participants to see the connection between the learning opportunities and what they need to learn (Corbalan et al., 2009; Killingsworth et al., 2015). Third, the presence of feedback allows participants to improve their self-regulation (Corbalan et al., 2009; Erhel & Jamet, 2013; Stobbeleir, Ashford, & Buyens, 2011). An identified self-regulation tactic is feedback-seeking behavior, where individuals search for information about their performance (Stobbeleir et al., 2011). Fourth, one of the most important factors of feedback is providing the participant with information regarding correcting inaccurate information and the correctness of their responses (Cameron & Dwyer, 2005; M. A. Evans, Pruett, Chang, & Nino, 2014). Finally, when a training curriculum provides feedback on performance, the acquired skills are better ingrained and resilient to skill decay (Kang, McDermott, & Roediger, 2007; Stefanidis, Korndorffer, Markley, Sierra, & Scott, 2006).

Within game environments, various ways of feedback can be provided, for example, points detracted or awarded based on performance, cues alerting incorrect or correct responses, and how a participant scored comparing others (Burgers et al., 2015; Ricci, Salas, & Cannon-Bowers, 1996).

2.2 Types of feedback

To increase learning in educational games, many types of feedback are available, depending on their complexity, specificity, length, and timing (Erhel & Jamet, 2013; Hattie & Timperley, 2007). Multiple studies about feedback in digital game-based learning focused on formative and summative feedback

(Lookadoo et al., 2017; Narciss et al., 2014; Serge et al., 2013; Van Mourik, 2020). Summative feedback provides learners with knowledge of their performance after a test, a task, or a set of tasks (e.g., grade, pass or fail, number of errors) (Narciss et al., 2014; Van Mourik, 2020). However, this type of verification information is the only knowledge learners receive with summative feedback (Serge et al., 2013). Formative feedback provides learners with information about their current behavior or thinking, to improve their learning (Narciss et al., 2014; Serge et al., 2013; Shute, 2008). One advantage of formative and summative feedback is the possibility to make it adaptive to the learner's needs, in this way the feedback becomes personalized and can, therefore, be directly bound to the learner's personal context (Narciss et al., 2014; Shute, 2008; Van Mourik, 2020).

2.3 Feedback specificity

Learning how to play a game can be quite challenging and may lead to an increase in player's cognitive load, which can affect the learning environment negatively (Serge et al., 2013). When the players encounter a new task without specific guidance or instruction the training can be experienced as overwhelming, due to the lack of clear direction on how to correctly perform the new task (Serge et al., 2013). Billings (2012); Johnson, Bailey, and Van Buskirk (2017); Serge et al. (2013) stated that performance improves when novice learners are provided with detailed feedback. However, when the mastery of the player increases, the player could get distracted by feedback, if it provides information that they already know. When a player's skills increase the level of detailed feedback should decrease (Serge et al., 2013). Changes in the feedback specificity and content can occur when a person learns, therefore, it is important to provide the learners with the right feedback at the right times to improve performance (Billings, 2012).

According to Serge et al. (2013); Shute (2008) the level of feedback specificity can vary with formative feedback, which indicates the level of information present in the feedback message. The level of specificity of formative feedback can range from vague and general, to detailed and specific (Serge et al., 2013). More detailed information on learners' actions and errors are provided when the level of feedback specificity increases (Goodman, Wood, & Chen, 2011; Serge et al., 2013). Detailed feedback provide the learner with explicit and clear instructions on how to perform a certain task, or how to correct specific errors in their gameplay (Billings, 2012; Shute, 2008). General or less specific feedback is not as directive as detailed feedback, the learners are provided with broad and conceptual suggestions such as hints (Billings, 2012; Shute, 2008). For example, with general feedback learners are informed that they made errors, whereas detailed and specific feedback also tells which actions were incorrect and correct (Goodman et al., 2011).

2.4 Feedback timing

The timing of the feedback message is also an important factor (Johnson et al., 2017). Regarding timing, the feedback message is usually delivered by either a delayed or an immediate approach (Billings, 2012; Hattie & Timperley, 2007; Smits, Boon, Sluijsmans, & van Gog, 2008).

Immediately feedback can be defined as the guidance immediately after completing a test, task, or problem (Billings, 2012; Shute, 2008), this type of feedback is beneficial for motor learning tasks and drill-and-practice tasks (Smits et al., 2008). For complex cognitive tasks, immediate feedback after a whole task is more desirable, the learning process is then not interrupted which gives the learner the possibility to understand the solution as a whole (Smits et al., 2008).

Delayed feedback can be defined as feedback that is provided after each training session or a series of tasks, the time of delay can vary between minutes, seconds, or even days (Billings, 2012; Clariana, Wagner, & Murphy, 2000; Hattie & Timperley, 2007; Smits et al., 2008). Delayed feedback is especially effective for real-time and complex tasks (Billings, 2012).

As mentioned above, besides the feedback timing, there is also a variance in the number of intervening elements (i.e. whole tasks or solutions steps) between immediate and delated feedback (Smits et al., 2008). Regarding the timing of feedback, there are mixed results and conflicting perspectives in literature (Johnson et al., 2017; Shute, 2008). According to Billings (2012); Bolton (2006) delayed feedback is a better alternative than immediate feedback because delayed feedback does not interrupt a task, which makes the game or scenario a better resemble of the real world. However, Johnson et al. (2017); Serge et al. (2013) stated that by providing learners with immediate feedback during serious games, the extraneous cognitive load for novice players reduces. When novices learn new procedures in serious games, the use of immediate feedback should be considered (Johnson et al., 2017).

2.5 Meaning feedback strategy

According to Narciss (2012); Narciss et al. (2014) a feedback strategy can be defined as a coordinated plan integrating decisive and clear statements that should specify the following aspects regarding the learning process:

The first aspect is the function and scope of the feedback, which can be defined as the purposes or goals the feedback serves. Feedback can have many different functions because it can affect the learning process at various levels (Narciss, 2012; Narciss et al., 2014). Based upon multiple feedback models the feedback functions can be classified on a cognitive, metacognitive, and motivational level (Narciss, 2013). According to Narciss (2013), feedback on a cognitive level informs, completes, corrects,

specifies, and restructures. Feedback on a cognitive level helps to recognize errors, acquire lacking knowledge, correct incorrect knowledge and associations, and specify inaccurate knowledge (Narciss, 2013). Feedback on a meta-cognitive level informs, completes, corrects, and guides. This type of feedback helps to recognize incorrect strategies, correct the incorrect strategies, attract attention to strategies, and acquire the missing strategies (Narciss, 2013). Feedback on a motivational level decreases task difficulty, increases incentive, associates success to effort, increases the probability of success, and increases the probability of positive perceptions of competence (Narciss, 2013).

The second aspect is the schedule and timing of the feedback, this means identifying the learning process events that trigger feedback messages. The third aspect is the content of the feedback, meaning what information the feedback should include. The fourth aspect is the conditions of the feedback, meaning under which individual and situational conditions the feedback should be provided. The last aspect is the presentation of the feedback, meaning in which modes and form the feedback is presented to the learner.

2.6 Feedback models

To design and examine adaptive feedback strategies a multidimensional view of feedback is needed (Narciss 2013). There are at least three facets of feedback that determine the quality and nature of feedback (Narciss et al., 2014). These facets are the functions of feedback, the contents of feedback, and the presentation of feedback contents (Narciss & Huth, 2004; Narciss et al., 2014). Taking these facets into account, the individual and situational conditions under which the feedback is provided need to be defined (Narciss et al., 2014). Multiple studies designed empirically and theoretically based feedback models (frameworks), that can be used to help design formative feedback (Shute, 2008). Because the main goal of this study is to increase participants' performance in the Space Fortress game, two feedback models that showed positive effects on achievement were used for the design of a personalized feedback strategy.

Narciss and Huth (2004) created a model for the design of formative feedback. Both Narciss and Huth (2004) and Shute (2008) stated that adapting the presentation, content, and function format of the feedback message should be driven by considerations of the learner characteristics and instructional goals, to maximize the informative value of the feedback. According to this model, the characteristics of the learner consist of three elements: prior knowledge, abilities, and skills such as metacognitive skills and content knowledge; learning goals and objectives; academic motivation (e.g. self-efficacy, meta-motivational skills) (Narciss & Huth, 2004; Narciss et al., 2014; Shute, 2008). The instructional goals also consist of three elements: the learning tasks, the particular instructional objectives, and errors and

obstacles (Narciss & Huth, 2004; Narciss et al., 2014; Shute, 2008). The impact of the feedback on motivation and learning, which was based on this feedback model, was examined in multiple studies. These studies showed that systematically designed formative feedback has positive effects on motivation and achievement (Shute, 2008).

Mason and Bruning (2001) created a feedback model based on research that examined levels of elaboration and multiple types of feedback in relation to prior knowledge, achievement level, the timing of feedback, and task complexity. According to this model, taking into account the achievement level of the student (participant) and the nature of the learning task, are the first steps in designing effective feedback. Based on these two factors, the most effective timing of the feedback can be determined (Mason & Bruning, 2001; Shute, 2008). For example, lower ability students likely have a more limited knowledge base, which makes it harder to self-correct errors and process current information, therefore, low achieving participants may benefit more from immediate feedback (Mason & Bruning, 2001). When the timing is determined, the following step in the model is to consider the students' level of prior knowledge, to implement the most effective type of elaboration or verification (Mason & Bruning, 2001). Finally, based on the mentioned variables in this feedback model the right type of feedback can be determined (Mason & Bruning, 2001; Shute, 2008).

The feedback model of Mason and Bruning (2001) can provide insights in the right timing and type of feedback, based on the participants' prior knowledge, task, and achievement level. The model of Narciss and Huth (2004) is made to design formative feedback and provides insights into which factors interact with feedback to influence learning.

2.7 Learner characteristics and the effect on feedback efficiency

In the sciences of learning and cognition, the concept of learner characteristics is used to designate a target group of learners and define aspects of their social, academic, cognitive, or personal self that can influence what and how they learn (Drachsler & Kirschner, 2012; Van Mourik, 2020). According to Drachsler and Kirschner (2012); Narciss et al. (2014) there are often large differences between learners and their characteristics, including motivational, prior knowledge, affective state, meta-cognitive skills, or learning styles and strategies. These differences have an impact on the degree of guidance and support of the learning process and the structure of the instruction (Drachsler & Kirschner, 2012). The many individual factors that result from these learning characteristics, can influence the way how feedback is processed by each learner and can support the design of personalized feedback strategies (Narciss et al., 2014; Sedrakyan, Malmberg, Verbert, Järvelä, & Kirschner, 2020). By taking the characteristics of the learner into

account, tailored feedback messages for a category of learners or an individual learner can be created (Narciss et al., 2014).

According to Drachsler and Kirschner (2012); Narciss et al. (2014); Van Mourik (2020); Vandewaetere, Desmet, and Clarebout (2011), four categories of learner characteristics can be defined: (1) Demographic characteristics, such as gender, personality, age, language. (2) Professional characteristics, refer to attitudes, knowledge, and competences related to the task, for example, achieved scores of the learner. (3) Conditional characteristics, generally refer to motivation, self-efficacy, and meta-cognitive abilities, these characteristics tend to have an impact on the learning process but are most of the time not a part of the learning objectives. (4) Contextual characteristics, such as time pressure, external events like stress, and distractions.

By taking into account the characteristics of the learners it is expected that more effective, motivating, and/or efficient learning strategies can be designed (Drachsler & Kirschner, 2012; Van Mourik, 2020). To design an effective, motivating, and efficient personalized feedback strategy, this study took multiple characteristics of the participants into account.

3. Current study

Based on the wishes of NLR and the presented theoretical framework, the following research question was formulated:

What are the characteristics of a personalized feedback strategy aimed at improving participants' performance in the Space Fortress game?

To answer this research question, three sub-questions were formulated. As stated in literature, the presentation, content, and function format of the feedback message should be driven by considerations of the learner characteristics and instructional goals, to maximize the informative value of the feedback (Narciss & Huth, 2004; Shute, 2008). By taking the characteristics of the learner into account, tailored feedback messages for a category of learners or an individual learner can be created (Narciss et al., 2014).

Therefore, the sub-questions of this research were:

1) Which individual conditions are relevant for the design of personalized feedback?
2) What clusters of participants emerge based on performance scores and individual learner characteristics?
3) Which learning and performance indicators are relevant for the design of personalized feedback?

4. Methods

4.1 Design

For this study, an applied and correlational research design was used, to provide clear insights what the content, function, and presentation of personalized feedback for the Space Fortress game should be. This was based on the individual factors, learning indicators and, performance indicators that affected the participants' performance in the Space Fortress game.

As part of finding out which learning indicators, performance indicators, and individual factors can be used for creating personalized feedback, correlational research was conducted on the performance and questionnaire data of the participants. This type of research helped to recognize patterns, relationships, and trends in the qualitative and quantitative data from the Space Fortress game, this contributed to recognize which factors and indicators affected participants' performances of the Space Fortress game. Therefore, the performance scores functioned as the dependent variables, and the performance indicators, learning indicators, and individual factors functioned as the independent variables.

Besides, the participants were closely observed and asked to evaluate their gameplay and their experiences. A thematic analysis for this qualitative data was used, to get a broader view concerning the indicators and factors that played a role in the performance scores.

4.2 Participants

The samples of this study consisted of 10 participants (8 male; 2 female) with a mean age of 21,7 years (*SD* = 2.87) and 73 participants who already played the game. The sample of 10 participants were asked to participate in the SF-AIS training while being observed by the researcher. The criteria of inclusion were to be over 18 years old, to master the English language, and to be unknown with the Space Fortress game. Because participants' performed the SF-AIS training from home, an additional requirement was to have access to a laptop or computer, connected to the internet, and with the right specifications to run the Space Fortress game smoothly.

The background of the participants was not diverse, each participant originated from the Netherlands. Most participants were master (20%) or bachelor (50%) students at the University of Twente. The remaining 30% of the participants were university of applied sciences students. The most frequent studies among the participants were industrial engineering & management (20%) and technical computer science (20%). The other participants studied human resource management, civil engineering, chemical science & engineering, mechanical engineering, biomedical technology, and pedagogical management.

The participants were invited for this study in person and participated voluntarily. The participants gave their consent via a consent form, this consent form can be found in Appendix A.

Initially, this research also wanted to take the participants that already played the game into account. However, two problems occurred. The first problem was that the number of dropouts was so high that the remaining number of participants was scarce. Before the start of this research, 73 people started with the SF-AIS training. Of these 73 people, 31 people quit playing after only completing the first training part. 22 people quit playing after the second training part and 7 people quit playing between training part 3 and the practice sessions. Therefore, the data of the dropouts were too incomplete to analyze and combine with the data of the 10 participants. The second problem was that the participants in this research had a custom-built SF-AIS, where the mandatory waiting time of 3 hours between the sessions was removed, which meant that they performed a different training. Therefore, the data could not be combined with the participants that already performed the SF-AIS training.

4.3 Instrumentation

4.3.1 The Space Fortress game

Space Fortress (SF) is a video game in which the participant is in control of a spaceship that navigates through space. The participant's objective is to gain as many points as possible by destroying the Space Fortress in the center of the screen. To destroy it, the fortress needs to be made vulnerable by hitting it 10 times, the number of hits against the fortress is displayed on the vulnerability counter, which can be found next to the fortress. When the vulnerability counter reaches 10, the participant can destroy the fortress with 2 quick shots. However, the fortress also fires at the spaceship which can cause damage, the spaceship can be protected by dodging those shots. Besides, while flying, the participant can earn points if the spaceship stays within the green hexagon area, does not bump into the fortress, and avoids hyperspace. When the spaceship is flying outside the green hexagon area, it easily leaves the screen completely. This will engage hyperspace, which means that the ship will be teleported to the opposite position on the screen.

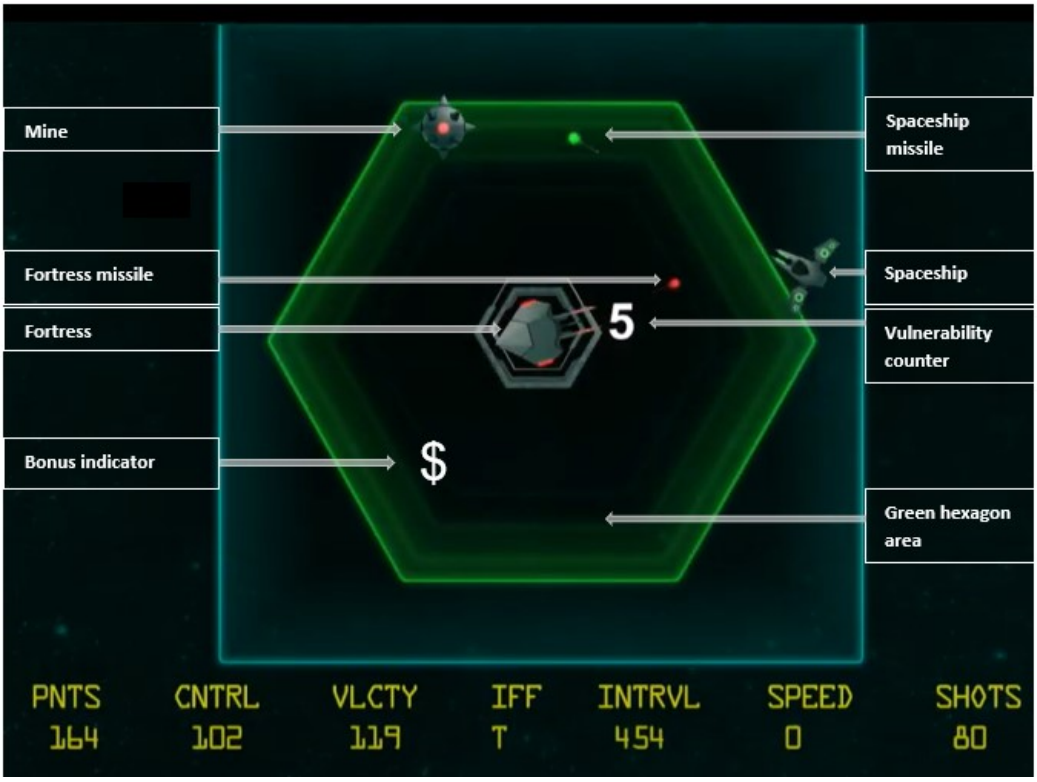
The spaceship also needs to be defended against two types of mines that appear at set intervals, these mines can be identified as either 'foe mine' or 'friendly mine' by monitoring the letter that appears under the label IFF (Identify Friend or Foe). Before the game starts, the participant will be briefed which three letters indicate a foe mine, any other letter indicate a friendly mine (Mané & Donchin, 1989). Friendly mines need to be activated by simply shooting them from nearby. Foe mines are harder to

destroy, the participant needs to identify this mine by pressing the J key twice with an interval of 250-400 msec, after that the participant can eliminate the mine by shooting it.

At the beginning of each game, the participant is given a supply of 100 missiles, which can be used for firing at the mines and fortress. However, once the participant is out of missiles points are subtracted for every missile that is fired. Therefore, the participant can control the number of available missiles, by collecting a bonus. The participant can collect this particular bonus of 50 more missiles when the $ sign appears twice in a row on the screen, if more missiles are needed, the K key needs to be pressed when $ appears the second time. However, if there is no need for extra missiles the participant can decide to collect 100 free points instead, therefore the L key needs to be pressed when $ appears for the second time.

Figure 1

Screenshot of the Space Fortress game



After each game, the participant receives an overview of their obtained scores (see Figure 2). The points (PNTS) score increases by shooting and destroying the fortress and collecting the bonus of 100 free points, reduces by being shot or destroyed. The control (CNTRL) score increases by flying within the

green hexagon area and reduces by bumping into the fortress and/or ending up in hyperspace. The velocity (VLCTY) score increases by minimizing high-speed movements and reduces by flying at high speed. The speed (SPEED) score increases by correct and quick responses on mines and decreases by being hit by mines. The total score combines PNTS, CNTRL, VLCTY, and SPEED. These scores are the only feedback which the game provided. Besides, no further explanation about these scores is given.

Figure 2

Overview of the scores shown after each game

4.3.2 Space Fortress Adaptive Instructional System (SF-AIS)

The Space Fortress game was originally developed in the 1980s by Mané and Donchin to study complex skill acquisition (Mané & Donchin, 1989; Stern et al., 2011). Besides its usefulness as a task for skill acquisition research, the SF game is also a useful research tool in studies on human learning, cognitive psychology, and machine learning (Van der Pal & Toubman, 2020).

To facilitate the NLR skill decay study, NLR built a reconstructed web-based version of the Space Fortress game, the so-called Space Fortress Adaptive Instructional System (SF-AIS) (Van der Pal & Toubman, 2020). The SF-AIS can be used as a measurement and training instrument, within this NLR project, both types of instruments are used. The participants learn how to play the Space Fortress game as they receive multiple learning tasks, intending to increase their skills. The performance data of these participants are then collected, to make a predictive model of skill retention (Van der Pal & Toubman, 2020). The SF-AIS can assign several types of tasks (e.g. learning paths, didactical actions, Space Fortress games with varying configurations, analyses) to the participants (Van der Pal & Toubman, 2020).

In this study, the custom-built SF-AIS was used to learn 10 participants the SF game and collect their performance data. Participants were given a reconstructed version of the SF-AIS, which was quite similar to the standard SF learning path that is used in the NLR skill decay study. In this reconstructed version, the mandatory waiting time of 3 hours between the sessions was removed, to play multiple sessions in a row. Besides, this reconstructed version consisted of the full initial training with various game elements, didactical actions were used to provide the participants with small learning objectives and a description of the various game elements. When the participants finished the initial training a series of full games were presented to measure their personal baseline game score, this baseline score indicated if the participants reached the minimally required score to continue to the next phase.

4.3.3 Learner measurement questionnaire

The learner measurement questionnaire of NLR was used to obtain participants' demographics and to assess participants' attitudes towards the initial training phase of the SF-AIS. Participants were asked about their gender, age, in which country they live, the highest level of education, hours a day spending behind the screen, gaming frequency, and profession (see Appendix B).

Participants' gaming experience was measured with the use of the learner measurement questionnaire, designed by NLR. Participants were asked how often they played platform, vector arcade, and MOBA games over the past half-year. MOBA games (e.g. League of Legends) are about a war between two teams, the player needs to control one of these teams strategically. Vector arcade games (e.g. Pac-Man) contain minimal graphics and have especially been played in the '80s. Platform games (e.g. Super Mario Bros) are games in which the player controls a character that encounters enemies and points, and has to jump from platform to platform. The responses on this questionnaire ranged between five options, (1) never, (2) 1x a month or less, (3) 1-4x a month, (4) >1x a week, and (5) prefer not to say. This outcome variable was used to examine if gaming frequency correlated with participants' performance in the game and to identify the different categories (group) of participants.

To assess participant's attitudes towards the training, the participants received eight statements that were measured on a 7-point Likert scale, ranged from (1) not at all true to (7) very true. The participants' needed to indicate to what extent they agree with statements such as ''I received enough feedback about my gameplay'' and ''I have mastered the game up to a good level''.

4.3.4 Situational Motivation Scale

The Situational Motivation Scale (SIMS) of Guay, Vallerand, and Blanchard (2000) was used as a measure of motivation. The SIMS consisted of intrinsic motivation, amotivation, identified regulation, and an external regulation scale. According to Guay et al. (2000) intrinsic motivation is when participants are being engaged for their own sake, for satisfaction, and the pleasure derived from performing the tasks. Unmotivated (amotivation) participants experience a lack of contingency between their outcomes and behaviors, they often experience feelings of incompetence. Identified regulation occurs when behavior is seen as being chosen by oneself. External regulation occurs when participants' behavior is regulated to avoid negative consequences or by rewards (Guay et al., 2000).

This scale consisted of 16 items that were scored on a 7-point Likert-scale, this ranged from (1) not at all true to (7) very true. Based on the question 'why are you currently engaged in this activity?', the participants needed to fill in the items of the SIMS. Examples of the amotivation scale items are: (1) I do this activity but I am not sure if it is worth it and (2) I don't know, I don't see what this activity brings me. Examples of the intrinsic motivation scale items are: (1) Because I think that this activity is interesting, and (2) Because I feel good when doing this activity. Examples of the external regulation scale are: (1) Because it is something that I have to do, and (2) Because I feel that I have to do it. Examples of the identified regulation scale are: (1) Because I think that this activity is good for me, and (2) By personal decision (Guay et al., 2000). This outcome variable was used to examine if motivation correlated with participants' performances in the game.

4.3.5 New General Self-Efficacy Scale

To assess the self-efficacy of the participants, the New General Self-Efficacy scale (NGSE) of Chen, Gully, and Eden (2001) was used. This scale consisted of eight items that were scored on a 5-point Likert scale, ranged from (1) strongly disagree to (5) strongly agree. Examples of items are: (1) In general, I think that I can obtain outcomes that are important to me, and (2) I will be able to successfully overcome many challenges (Chen et al., 2001). This outcome variable was used to examine if self-efficacy correlated with participants' performances in the game.

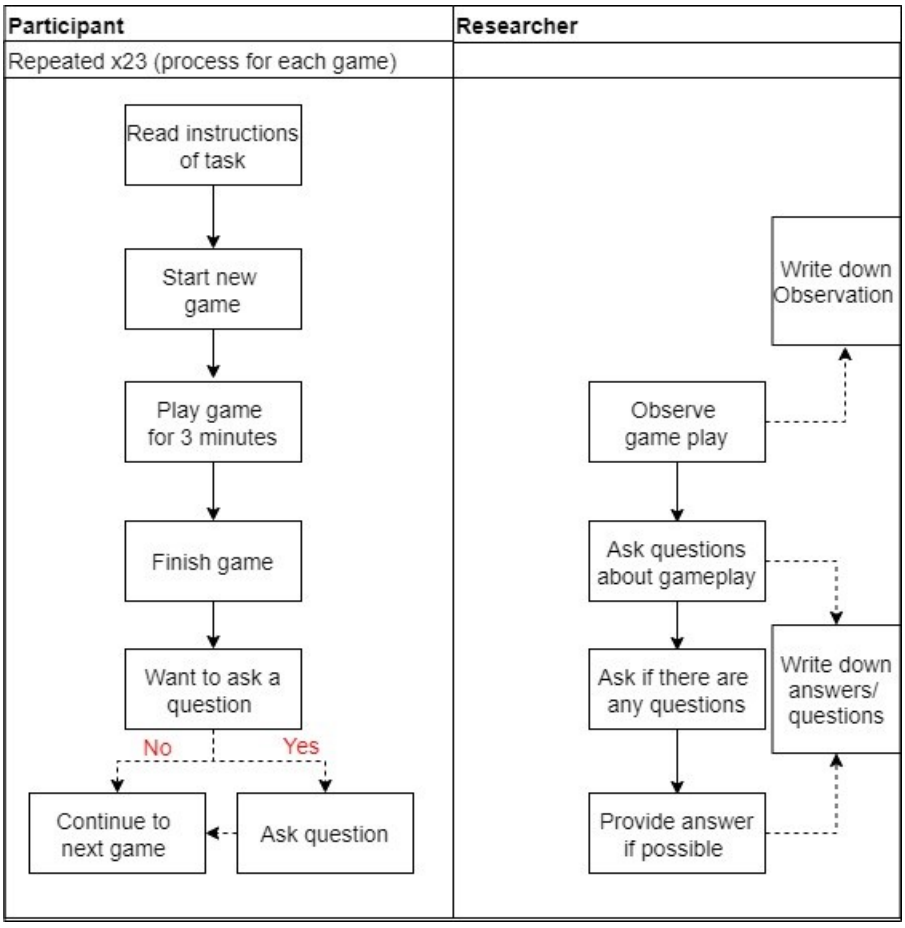
4.4 Procedure

Because people were involved in this study, the ethics commission of the University of Twente was asked for official permission to work with the data. After the permission was granted, the participants received information about the purpose of the study and were asked to sign a consent form in real life. The 10 participants performed the initial training online, in the custom-built Space Fortress Adaptive Instructional System of NLR. This training consisted of two individual sessions that included three training parts and one practice session, both training sessions took approximately one hour to complete. The first training session consisted of two training parts, training part 1 included six games, and training part 2 included five games, each game had a duration of 3 minutes. The second training session consisted of one training part and one practice session, training part 3 consisted of four games, and the practice session consisted of eight full games of the game. See Figure 3 for a schematic overview of the data collection process, this process was repeated for each game that the participants played.

Two individual training sessions were provided per participant with one hour apart from each other. Within each session, the participants played the Space Fortress game and answered the self-efficacy, motivation, and learner measurement questionnaires, to examine participants' learner characteristics and experiences of the Space Fortress game. Due to COVID-19, the participants participated from home and used their laptop or computer, during both sessions the participants were closely observed by the researcher. After the completion of each learning task in the game, the researcher asked the participants ‘how did it go?’, ‘did you run into something?’, and ‘do you have any questions?’. After the participants completed the two sessions they were thanked for their participation in this study.

Figure 3

Schematic overview of the data collection process



4.5 Analysis

4.5.1 Situational Motivation Scale

The responses of the participants on this questionnaire ranged between seven options, (1) strongly disagree, (2) disagree, (3) somewhat disagree, (4) neutral, (5) somewhat agree, (6) agree, (7) strongly agree. The calculated scores ranged from 1-7 respectively and measured participants' motivation.

To answer the first sub-question, the first step was to conduct a descriptive analysis and analyse the data of the motivation variable, using IBM SPSS Statistics (Version 25) predictive analytics software. A higher score on the intrinsic motivation and identified regulation scales indicated that the participant had a positive intrinsic motivation and identified regulation, whereas a lower score indicated that the participant did not experience a positive intrinsic motivation and identified regulation. For the amotivation and the external regulation scales, the interpretation of the scores was exactly the opposite.

4.5.2 New General Self-Efficacy Scale

The responses of the participants on this questionnaire ranged between five options, (1) strongly disagree, (2) disagree, (3) neutral, (4) agree, (5) strongly agree. The calculated scores ranged from 1-5 respectively and measured participants' self-efficacy. To answer the first sub-question, the same analyses and software as the Situational Motivation Scale (motivation variable) were used. A higher score on this questionnaire indicated that the participant had a high self-efficacy, whereas a lower score indicated the participant had a low self-efficacy.

4.5.3 Gaming experience

The responses of the participants on this questionnaire ranged between five options, (1) never, (2) 1x a month or less, (3) 1-4x a month, (4) >1x a week, and (5) prefer not to say. To answer the first sub-question, an exploratory and descriptive data analysis was conducted to give insights into the gaming frequency of the participants, using IBM SPSS Statistics (Version 25) predictive analytics software.

4.5.4 Game performance

To measure participant's performance in the game, the sub-scores of the variables: points, speed, control, and velocity were obtained. Each sub-score had a different range of scores. Practically speaking, each variable has a minimum and a maximum score, however, NLR did not found a reliable way to calculate those scores yet. From the control and velocity sub-scores, the maximum score is known, the other minimum and maximum scores are still unknown. Therefore, the following assumptions of the score ranges are made based on the performance scores of the participants.

The control sub-score ranged between -700 and 1080, whereas 1080 is the maximum score that could be achieved. The points sub-score ranged between -2000 and 5000. The speed sub-score ranged between -400 and 600. The velocity sub-score ranged between -100 and 1260, whereas 1260 is the maximum score that could be achieved.

From each variable the median was calculated, which functioned as an indicator of participants' game performance. After calculating the medians of the game performances, the relationship between the individual variables (motivation, self-efficacy, gaming experience) and the game performances was calculated, by conducting a Pearson correlation analysis. To find out if the variables were normally distributed, histograms were created and assessed by a visual inspection. To calculate the mean, create the histograms, and perform the correlation test, IBM SPSS Statistics (Version 25) predictive analytics software was used.

The strength of the correlations were described with the use of Evans' classification. According to Evans (1996), the strength of the absolute value of *r* can verbally be described as: .00-.19 ''very weak'', .20-.39 ''weak'', .40-.59 ''moderate'', .60-.79 ''strong'', and .80-1.0 ''very strong''.

This outcome variable was used to examine the differences in participants' game performances to identify groups with common traits, how these game performances correlated with the individual conditions of the participants, and which learning and performance indicators were relevant.

4.5.5 Exploratory data analysis

To identify which groups exist among the participants, an exploratory data analysis was conducted. Graphs of performance scores were created using Microsoft Excel (2016). Besides the outcomes of the multiple questionnaires measuring the learners' characteristics, were analyzed to find patterns. These graphs and questionnaire outcomes were used to identify groups that are characterized by common traits in participant's learner characteristics and performances (e.g. high achievers in all subjects, participants that excel in certain learning tasks but fail in others).

4.5.6 Thematic analysis

While participating in this study, the participants were closely observed by the researcher. During the observations, the researcher focused on participants' reactions and feelings while playing the game (e.g. frustrated, engaged, distracted) and the possible actions and tasks that caused them, the errors made, and incorrect strategies the participants applied. Because the researcher observed the participants, a possible threat was that the participants changed their behaviors because of it, which could have led to different outcomes as participants played the game on their own. During the observations, the researcher wrote everything down in a Word file. After the completion of each learning task in the game, the researcher asked the participants 'how did it go?', 'did you run into something?', and 'do you have any questions?'. The answers to these questions and the questions of the participants themself were written down as well. An inductive coding approach was used to determine the themes, which means that the data determined the themes. Besides a semantic approach was used, which means that the explicit content of the data was analyzed.

To conduct a thematic analysis, the six phases of analysis from Braun and Clarke (2006) were used. Braun and Clarke (2006) defined the following six phases of thematic analysis: (1) *Familiarization*, noting down initial ideas, reading and re-reading the data, and transcribing data. (2) *Generating initial codes*, coding the data in a systematic way across the entire data set, collating relevant data for each code. (3) *Generating themes*, identify patterns in the codes to come up with relevant themes, combining several

codes into one theme. (4) *Reviewing themes,* check if the themes work in relation to the entire data set and the coded extract. (5) *Defining and naming themes,* analysis to define the specifics of each theme, naming and defining each theme. (6) *Producing the report,* writing the analysis of the data, how it relates to the research question and literature.

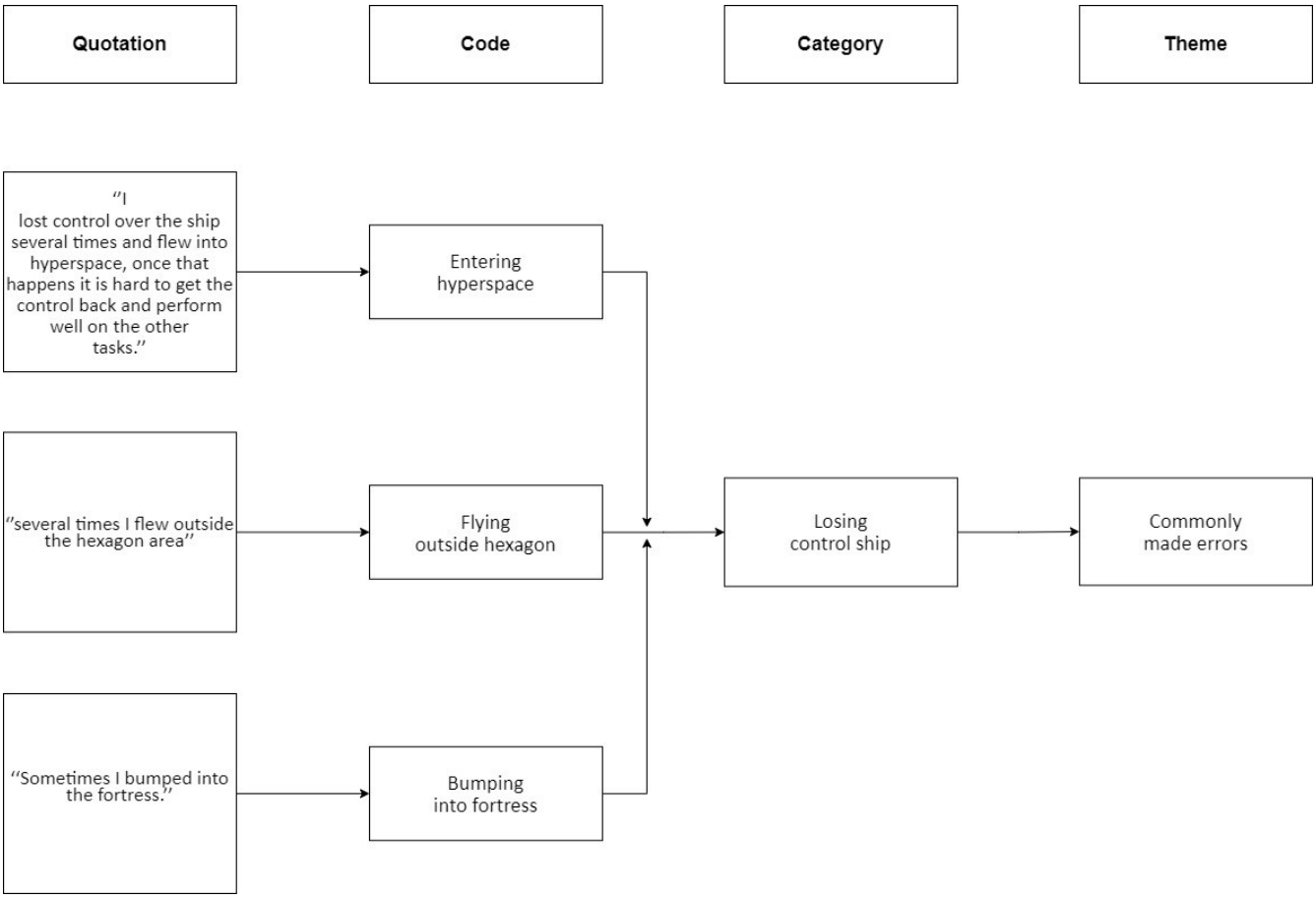
The ATLAS.ti 9 software package was used to analyse the data. The obtained information of the participants was divided into smaller parts, using several core codes. Participant's experiences were divided under 'experienced difficulties and challenges' and 'experienced strengths', the questions of the participants were coded as 'questions about the game and task', and the observations were divided under 'observed difficulties and challenges', 'observed strengths', and 'observed additional points'. An example quotation from the code 'experienced difficulties and challenges' is ''I find It hard to focus on shooting and flying at the same time, I also find it quite unclear when the fort is being destroyed''. A total of 21 codes were created, these codes can be found in Appendix E and F. The information of the participants was first coded in ATLAS.ti 9. After coding, an overview of the results for each code was made with the use of Microsoft word. The codes were then combined into categories, like 'frustrated', 'stressed', and 'bored' into the category emotions. All the quotations, codes, and categories were then compared to define the overarching themes, an overview of the whole thematic analysis can be found in Appendix E. An example of a theme is commonly made errors, multiple categories and codes are combined to establish this theme. Figure 4 illustrates how multiple quotations and codes lead to one category and theme.

To increase the internal validity of these qualitative results, participant validation was conducted. According to Birt, Scott, Cavers, Campbell, and Walter (2016) a participant validation is the method of returning analyzed data or an interview to a participant to confirm and check the results. In this study, two participants were asked to check and confirm the qualitative results.

This analysis was used to answer the second sub-question, which learning and performance indicators were relevant for the design of personalized feedback.

Figure 4

Example of multiple quotations and codes leading to one category and theme



5. Results

5.1 Descriptive statistics

The mean scores on the control, points, speed, and velocity variables for each training part are presented in Tables 1 and 2.

Table 1

Descriptive statistics of performance scores Training session 1

	Training session 1							
	Training part 1 (*n* = 10)				Training part 2 (*n* = 10)			
Sub-score	*Min*	*Max*	*M*	*SD*	*Min*	*Max*	*M*	*SD*
Control score[a]	201	855	587.10	237.19	316	1031	808.00	211.20
Points score[b]	159.00	2063.0	908.68	730.31	-255	1826	779.00	731.30
Speed score[c]	-400	378	-57.10	234.41	-147	320	89.70	165.03
Velocity score[d]	768	1115	946.00	133.01	1151	1252	1210.1	32.47

[a] Scale between -700 and 1080
[b] Scale between -2000 and 5000
[c] Scale between -400 and 600
[d] Scale between -100 and 1260

Table 2

Descriptive statistics of performance scores Training session 2

	Training session 2							
	Training part 3 (*n* = 10)				Practice session (*n* = 10)			
Sub-score	*Min*	*Max*	*M*	*SD*	*Min*	*Max*	*M*	*SD*
Control score[a]	332	991	712.50	220.45	620	972	835.00	124.95
Points score[b]	-298	1366	566.50	566.35	-216	1684	873.10	625.55
Speed score[c]	-66	289	126.70	125.10	-55	231	135.10	82.33
Velocity score[d]	1048	1239	1151.3	64.28	1146	1257	1218.0	40.66

[a] Scale between -700 and 1080
[b] Scale between -2000 and 5000
[c] Scale between -400 and 600
[d] Scale between -100 and 1260

The mean scores on self-efficacy, motivation, and the four scales of the SIMS, intrinsic motivation, amotivation, identified regulation, and external regulation are presented in Table 3. The five-point Likert scale used for measuring self-efficacy was considered an interval scale. The interval length was 0.80, whereas five interval ranges were distinguished. From 1 to 1.8 (strongly disagree), from 1.81 to 2.60 (disagree), from 2.61 to 3.40 (neutral), from 3.41 to 4.20 (agree), and from 4.21 to 5 (strongly agree). Self-efficacy showed a mean of 3.81 which can verbally be interpreted as high. To conclude, on average the participants had a high self-efficacy.

The seven-point Likert scale used for measuring motivation was also considered an interval scale. The interval length was 0.86, whereas seven interval ranges were distinguished. From 1 to 1.86 (strongly disagree), from 1.87 to 2.71 (disagree), from 2.72 to 3.57 (somewhat disagree), from 3.58 to 4.43 (neutral), from 4.44 to 5.29 (somewhat agree), from 5.30 to 6.14 (agree), and from 6.15 to 7 (strongly agree). Intrinsic motivation showed an average mean of 4.05, which showed that on average participants had neither high nor low intrinsic motivation. Identified regulation showed an average mean of 3.15, which showed that on average participants had a moderately low identified regulation. Amotivation showed an average mean of 2.80, which showed that on average participants had a moderately low amotivation. External regulation showed an average mean of 1.63, which showed that on average participants had a very low external regulation. The motivation variable included intrinsic motivation, amotivation, identified regulation, and external regulation. The average mean of motivation was 2.91, which showed that participants had moderately low motivation.

Table 3

Descriptive statistics Self-efficacy and Motivation

	(n = 10)			
Variable	*Min*	*Max*	*M*	*SD*
Self-efficacy	3.00	4.25	3.81	.36
Intrinsic motivation	2.75	5.25	4.05	.95
Amotivation	1.00	4.50	2.80	1.11
Identified regulation	1.75	4.50	3.15	.74
External regulation	1.00	3.50	1.63	.94
Motivation	1.63	3.38	2.91	.51

The frequency distribution of gaming frequency is presented in Table 4. This frequency distribution showed that in the last six months, platformer games were the most played games among the participants, each participant played platformer games. Vectors arcade games were the least played games among the participants, 50 percent of the participants never played vector arcade games in the last six months, while the other 50 percent played 1x a month or less. The number of participants that played MOBA games was more equally divided, some participants never played MOBA games while others played this type of games weekly.

Table 4

Frequency distribution of gaming frequency over the last six months by type of game

	Platform games		Vector arcade games		MOBA games	
Gaming frequency last 6 months	*n*	%	*n*	%	*n*	%
Never			5	50	2	20
1x a month or less	6	60	5	50	4	40
1-4x a month or less	3	30			1	10
>1x a week	1	10			3	30

5.2 Statistical relationship between individual variables and game performances

To answer the first sub-question 'which individual conditions are relevant for the design of personalized feedback?', the statistical relationships between the individual variables and the game performances were calculated with the use of Pearson's correlation coefficient, see Table 5.

Table 5

Correlation table of the sub-scores on the full Space Fortress game (game performance), the variables motivation (SIMS), self-efficacy (NGSE), and gaming frequency.

	NGSE	SIMS	Velocity	Speed	Points	Control	Gaming frequency
NGSE		.626	.475	.089	.177	.368	.232
SIMS			.285	.207	.353	.171	.363
Velocity				.382	.803**	.509	.282
Speed					.654*	.676*	.703*
Points						.533	.590
Control							.210
Gaming frequency							

*. Correlation is significant at the 0.05 level (2-tailed).

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

In general, a strong positive correlation was found between the self-efficacy and motivation variables. Very weak, weak, and moderate positive correlations were found between self-efficacy, gaming frequency, and performance scores. The self-efficacy had a strong positive correlation with motivation r = .626, a positive moderate correlation with the velocity sub-score r = .475, a positive weak correlation with the control sub-score r = .368, and the gaming frequency r = .232, and a positive very weak correlation with the speed sub-score r = .089, and the points sub-score r = .177.

The motivation variable also had no strong correlations with the performance scores and the gaming frequency. Motivation had a positive weak correlation with the velocity sub-score r = .285, the speed sub-score r = .207, the points sub-score r = .353, and the gaming frequency r = .363, and a positive very weak correlation with the control sub-score r = .171.

Stronger correlations were found among the sub-score variables and between the sub-score variables, and gaming frequency. The velocity sub-score had a very strong positive correlation with the points sub-score r = .803, a moderate positive correlation with the control sub-score r = .509, a weak positive correlation with the speed sub-score r = .382, and the gaming frequency r = .282. The speed sub-score had a strong positive correlation with the points sub-score r = .654, the control sub-score r = .676, and the gaming frequency r = .703. The points sub-score had a positive moderate correlation with the

control sub-score *r* = .533, and the gaming frequency *r* = .590. The control sub-score had a weak positive correlation with gaming frequency *r* = .210.

5.3 Exploratory data analysis

To answer the second sub-question, 'What clusters of participants emerge based on performance scores and individual learner characteristics?'. Graphs and questionnaire outcomes were used to identify groups that are characterized by common traits in participant's learner characteristics and performances.

First, the total mean of each sub-score was calculated. These mean scores were based on the 10 participants and the 23 learning tasks. Table 6 shows an overview of the means of each sub-score and the total score.

Table 6

Total mean of sub-scores

Sub-scores	*M*
Control score	736
Points score	782
Speed score	74
Velocity score	1131
Total score	2696

These means were the average baseline score for each sub-score. Second, the mean scores of each participant were calculated. These outcomes were then compared with the total mean of the sub-scores. In this way, a distinguish between participants was made based on their performances. For each sub-score a graph was created, these graphs provided clear insight if participants scored above or below average. These graphs (Figure 5, 6, 7, 8, and 9) can be found in Appendix G.

Figure 5 shows the mean and average score on the control sub-score. Figure 5 indicated that participants 2, 6, 7, 8, and 9 scored above average on the control sub-score, and participants 1, 3, 4, 5, and 10 scored below average on the control sub-score.

Figure 6 shows the mean and average score on the points sub-score. Figure 6 indicated that participants 2, 3, 4, 7, and 9 scored above average on the points sub-score, and participants 1, 5, 6, 8, and 10 scored below average on the points sub-score.

Figure 7 shows the mean and average score on the speed sub-score. Figure 7 indicated that participants 2, 3, 7, 8, 9, and 10 scored above average on the speed sub-score, and participants 1, 4, 5, and 6 scored below average on the speed sub-score.

Figure 8 shows the mean and average score on the velocity sub-score. Figure 8 indicated that participants 2, 3, 4, 6, and 9 scored above average on the velocity sub-score, and participants 1, 5, 7, 8, and 10 scored below average on the velocity sub-score.

Figure 9 shows the mean and average score on the total score. Figure 9 indicated that participants 2, 3, 4, 7, and 9 scored above average on the total score, and participants 1, 5, 6, 8, and 10 scored below average on the total score.

Based on these performance scores three groups were identified. The first group was called the high-performing group and consisted of participants that scored above average on each sub-task or 3 of the 4 sub-tasks. The second group was called the middle-performing group and consisted of participants that had more varying scores for example two sub-tasks below average and the other two sub-tasks above average. The third group was called the low-performing group and consisted of participants that scored below average on each sub-task or 3 of the 4 sub-tasks.

5.3.1 High-performing group

Participants 2, 9, 3, and 7 were placed in the high-performing group. Participants 2 and 9 scored above average on each sub-task, participants 3 and 7 scored above average on 3 of the 4 sub-tasks. Besides the common traits in performances, the participants also had common traits in their learner characteristics.

The gaming frequency of the participants in this group was the highest among the participants, overall each participant in this group played a game between 1x a month or less and 1-4x a month. Besides the participants in this group had the highest average screen time, each participant spends more than 6 hours behind a screen every day.

Although the general motivation was moderately low, Table 7 indicated that the high-performing group had the ‘highest’ motivation among the participants. During the gameplay observations of these four participants, it became clear that they started to lose motivation when no further tasks were added to the gameplay. They were no longer challenged, which led to a decrease in their motivation. One participant tried to motivate himself by asking about the scores of the other participants, to make the game more competitive.

Table 7

Mean of motivation among the three different groups

	Group		
	High-performing (*n* = 4)	Middle-performing (*n* = 3)	Low-performing (*n* = 3)
Variable	*M*	*M*	*M*
Motivation	3.22	2.85	2.54

The participants in the high-performing group were also the most confident that they could play the Space Fortress game. The participants were asked to rate themselves on a scale from 1-10 how confident they were that they could play the Space Fortress game. Whereas 1 indicated 'not at all' and 10 'extremely confident'. On average the participants in this group rated themselves with an 8.5.

At the end of the training, the participants evaluated the Space Fortress game. The participants were asked to rate different aspects on a scale from 1- 7, three aspects were taken into account. The first aspect was if the participants found the game complicated (1) or easy (7), the average of the high-performing group was 4,25 which indicated that the participants found the game neither easy nor complicated. The second aspect was if the participants found the game easy to learn (1) or difficult to learn (7), the average of the high-performing group was 2.75 which indicated that the participants found the game moderate easy to learn. The third aspect was if the participants found the game boring (1) or exciting (7), the average of the high-performing group was 3.5 which indicated that the participants found the game moderate boring.

5.3.2 Middle-performing group

Participants 4, 6, and 8 were placed in the middle-performing group. Participant 4 scored above average on the points and velocity sub-score, and below average on the control and speed sub-score. Participant 6 scored above average on the control and velocity sub-score, and below average on the points and speed sub-score. Participant 8 scored above average on the control and speed sub-score, and below average on the points and velocity sub-score. Besides the common traits in performances, the participants also had common traits in their learner characteristics.

The gaming frequency and screen time of the participants in this group were not the highest nor the lowest among the participants, overall each participant in this group played a game between never and 1x a month or less. Besides the average screen time of the participants in this group was between 4-6 hours a day and more than 6 hours every day.

Table 7 indicated that the middle-performing group scored between the high-performing and low-performing on motivation. The score of 2.85 indicated that their motivation was moderately low.
During the gameplay observations of these three participants, it became clear that they struggled with some tasks and therefore made several mistakes during the game, this often led to frustration and a decrease in motivation.

On average the participants in this group rated themselves with a 7.7 on the question of how confident they were that they could play the Space Fortress game.

The average of the middle-performing group on the first evaluation aspect was 4, which indicated that the participants found the game neither easy nor complicated. The average on the second aspect was 3.33 which indicated that the participants found the game moderately easy to learn. The average on the third aspect was 3.66 which indicated that the participants found the game neither boring nor exciting.

5.3.3 Low-performing group

Participants 1, 10, and 5 were placed in the low-performing group. Participants 1 and 5 scored below average on each sub-task, participant 10 below average on 3 of the 4 sub-tasks. Besides the common traits in performances, the participants also had common traits in their learner characteristics.

The gaming frequency of the participants in this group was the lowest among the participants, overall each participant in this group played a game between never or 1x a month or less. Besides, the participants in this group had the lowest average screen time, between 2-4 hours and more than 6 hours a day.

Although the general motivation was moderately low, Table 7 indicated that the low-performing group had the 'lowest' motivation among the participants. The score of 2.54 indicated that their motivation was low. During the gameplay observations of these three participants, it became clear that they started to lose motivation because they often made mistakes, and struggled with handling the tasks. Besides, the participants in the low-performing group often did not understand how to improve their gameplay.

On average the participants in this group rated themselves with a 5.7 on the question of how confident they were that they could play the Space Fortress game.

The average of the low-performing group on the first evaluation aspect was 3, which indicated that the participants found the game moderately complicated. The average on the second aspect was 3.66

which indicated that the participants found the game not easy nor difficult to learn. The average on the third aspect was 5.66 which indicated that the participants found the game exciting to play.

5.4 Thematic analysis

To answer the third sub-question, 'Which learning and performance indicators are relevant for the design of personalized feedback?', the qualitative data from the participants were divided into three themes. The themes were identified as 'Commonly made errors', 'Typical incorrect strategies', and 'Potential learning problems'. These themes provided insights which errors were frequently made by the participants, which incorrect strategies participants used while playing the game, and the most frequent learning problems that the participants experienced.

5.4.1 Commonly made errors

Playing the Space Fortress game is a complex game, while playing this game each participant made several errors. Applying the thematic analysis technique with input from observations, experiences and questions, resulted in an overview of the commonly made errors the participants made while playing the Space Fortress game. An example of a commonly made error was losing the controls of the ship. Participants flew outside the green hexagon area, flew into hyperspace, and bumped into the Space Fortress. Table 8 shows the categories of this theme.

Table 8

Categories of commonly made errors

Categories
• Losing control ship • Hit by Space Fortress • Not destroying foe mines • Miss bonus opportunity

Losing control ship

As mentioned above the category losing control ship was based upon the codes 'entering hyperspace', 'flying outside hexagon', and 'bumping into the fortress'. The participants experienced that they often lost control over the ship:

''I bump into the fortress more often, if I know that the fortress would not shoot back.''

''Sometimes I bumped into the fortress.''

''several times I flew outside the hexagon area''

''I lost control over the ship several times and flew into hyperspace, once that happens it is hard to get the control back and perform well on the other tasks.''

''I need to get used to the controls of the ship, the W key is also really sensitive, if I press it too long you immediately fly fast.''

The participants also asked several questions about controlling the ship:

''Is there a consequence if I'm flying on the edge of the green hexagon?''

''Are points subtracted if I fly against the fortress?''

''How fast do I need to fly? This is unclear for me.''

''Will the ship always have this speed?''

The researcher observed the participants while playing the game and noticed that controlling the ship was a hard task to complete, especially at the start of the SF-AIS training. Multiple participants experienced stress or frustration once they lost control over the ship. The researcher noticed that once a participant lost the control over the ship, the other tasks were negatively influenced by this. The participants always focused to gain back the controls of the ship, before focussing on the other tasks again. Although losing control over the ship was a frequently made mistake, most participants made this mistake less often after playing multiple games.

Hit by Space Fortress

The category hit by Space Fortress was based upon the code 'not dodging shots', meaning when the participant was not able to dodge the shots of the Space Fortress with the spaceship. The participants experienced that they often got hit by the shots of the Space Fortress:

''I struggled with dodging the shots of the fortress, almost every shot hit me, this made me frustrated.''

''Not great, I was hit several times at the start of the game, once you have a bad start it is hard to play well again.''

''Not so great, I was hit a lot and died several times. I noticed a chain reaction, making multiple mistakes leads to making more mistakes.''

''I need to get used to the fact that the fortress is shooting faster, because of this I died twice.''

''Not that good, I did not notice that the fortress was shooting faster at me, I experienced more stress, this influenced my gameplay.''

Besides, the participants asked if the Space Fortress was going to shoot faster than it already did, if it was shooting faster, and if points were subtracted once the ship was hit by a shot:

''Will the fortress shoot faster than this in upcoming games?''

''Is the fortress shooting faster?''

''How many points are subtracted when the ship gets hit by a missile or mine?''

As participants progressed in the game, the Space Fortress was also shooting faster at them. However, this phenomenon was never told to the participants. The researcher noticed that the participants needed to get used to the new shooting speed of the Space Fortress, the ships were hit more often and therefore multiple participants died several times in the game. Most of the time, the participants were not fast enough with flying away from the shot, participants tend to stand still while they need to handle a mine for example. This often led to getting hit by the Space Fortress itself. The researcher could tell that most of the participants got nervous from the shots or frustrated because they got hit.

Not destroying foe mines

The category not destroying foe mines was based upon the codes 'wrong interval', 'forget letters', and 'hit by mine'. The commonly made mistakes were not being able to get the right interval time for the J key, forgot or forgot to remember the three letters indicating a foe mine, and being hit by a mine. The participants often experienced difficulties with handling the mines:

''I struggled to handle the foe mines, my average interval for the J key was 128, this is way too fast.''

''I experience some struggle with the J key, getting the right interval is hard.''

''I struggled with remembering the three letters indicating a foe mine, I did not understand that I needed to remember those, did not read the instructions clear enough.''

''I forgot the letters indicating a foe mine.''

''I struggled during this game, handling the mines was quite difficult. I got hit several times by the missiles of the fortress and the mines themselves. My main focus was on the mines and therefore I forgot to also focus on the other elements.''

''I got hit several times by the mines.''

''Sometimes I struggle to see if I defused the foe mines correctly.''

The participants also asked several questions about handling the mines:

''Are the letters indicating a foe mine only be showed at the begin of each game?''

''Is the & symbol also an indicator for mines?''

''Do you need to hit the foe mine once or twice to destroy it?''

''Are points also subtracted if you get hit by a friendly mine?''

The outcomes of the observations showed that handling the mines was for most participants the hardest task. Whereas controlling the ship showed clear improvements over time, handling the foe mines was a task where multiple participants struggled with until the end of the training. To destroy a foe mine the participant needs to accomplish multiple steps, for some people this was too hard. The most common mistakes were not getting the right interval for the J key, and being hit by the mine itself. A few participants forgot the three letters indicating a foe mine, however, most of the participants did not experience any difficulty remembering these letters.

Miss bonus opportunity

The category miss bonus opportunity was based upon the codes 'forget bonus, and 'collecting bonus'. The commonly made mistakes were forgot to collect a bonus and trying to collect a bonus at the wrong time (when a different symbol then $ appears on the screen, and when the $ did not appears twice in a row). The participants often experienced difficulties with the bonus opportunities:

''Sometimes I struggle with collecting the bonuses.''

''Sometimes I forget to collect a bonus.''

''Sometimes I forget to collect a bonus because there is no direct consequence visible when you continue shooting with zero bullets.''

There were some unclarities about the bonus task, this led to several questions from the participants:

''Can I also collect a bonus when the dollar sign appears three in a row (and I forgot to collect during the second time)?''

''If you missed the second dollar sign, will the change to collect a bonus be reset? Or will the game continue counting?''

''Do you still receive a bonus when you 'die', or is the counter starting on zero again?''

''So the dollar sign needs to appear on the screen, then disappear and if a second sign shows up then you can collect a bonus?''

''If I wrongfully collect a bonus, does the number of dollar signs restart as well?''

''Once you died, will the change to collect a bonus reset, or if you try to collect a bonus when another sign appears, will it reset then?''

The observations showed that people were merely focused on the bonus task, participants that experienced difficulties with the other tasks did not prioritize to collect a bonus. Participants that performed well on the other task, were more focused to collect a bonus. Forgetting how many times the dollar sign already occurred on the screen was a commonly made mistake. There was also a lot of unclarity about the exceptions of collecting a bonus, this led to multiple questions from different participants.

5.4.2 Common incorrect strategies

While playing the Space Fortress game multiple participants had incorrect strategies of playing the game. The observations, experiences, and questions provided an overview of the typical incorrect strategies the participants made while playing the Space Fortress game. An example of a commonly incorrect strategy was flying outside the hexagon on purpose. Participants flew outside the green hexagon area, to dodge the shots from the Space Fortress. Table 9 shows the categories of this theme.

Table 9

Categories of typical incorrect strategies

Categories
• Flying outside hexagon on purpose • Immediately collecting bonus • Indicate the type of mine by shooting at it • Staying at the starting point

Flying outside hexagon on purpose

The category flying outside hexagon on purpose was based upon the code 'flying wrong place'. The commonly incorrect strategy was flying outside the hexagon on purpose, to avoid being shot by the Space Fortress. The participants often used this strategy:

> *''I notice that I'm flying outside the hexagon and the square, to easily dodge the shots from the fortress.''*

Participants knew it was an incorrect strategy, and therefore wanted to know the consequences of it by asking questions:

> *''Is there a consequence if I'm flying on the edge of the green hexagon?''*
> *''Are there any consequences if you shoot outside the hexagon area?''*

Although not every participant used this incorrect strategy, the researcher observed that this strategy was frequently used intentionally and unintentionally by multiple participants. The participants were often better able to dodge the shots from the Space Fortress, however, because of this strategy the ships of the participants were more often hit by a mine. Because the mines occur in random places on the edge of the screen, people had less time to react to the mines, because they already flew closer to the edge. The participants clearly understood that this was an incorrect strategy, but did this anyway.

Immediately collecting bonus

The category immediately collecting bonus was based upon the code 'immediately collecting'. The commonly incorrect strategy was trying to collect a bonus instantly, when a dollar sign appears on the screen, not knowing how many times it occurred on the screen. Some participants used this strategy:

> *''Once I see a dollar sign I immediately try to collect a bonus, but most of the time that does not work.''*

As mentioned above, the participants asked several questions about collecting a bonus. However, no specific questions were asked regarding using this strategy.

While just two participants used this strategy on purpose, other participants used this strategy unintentionally. Participants frequently panicked when the dollar sign showed up on the screen, instead of counting the number of times the dollar sign appeared, multiple participants tried to instantly collect the bonus instead. Some participants claimed that they forgot how many times the dollar sign appeared, while others were too busy with the other tasks.

Indicate the type of mine by shooting at it

The category indicating the type of mine by shooting at it was based upon the code 'shooting at mine'. The commonly incorrect strategy was shooting at a mine to indicate if it is a foe mine or a friendly mine, instead of remembering the three letters indicating a foe mine. Some participants used this strategy:

> *''I noticed that I do not pay attention to IFF anymore, I just shoot at a mine and from there on see if it is foe or not.''*

As mentioned above, the participants asked several questions about handling the mines. However, no specific questions were asked regarding using this strategy.

While most participants used the letters and intervals to indicate and destroy the mines, a few participants decided to just shoot at the mine to indicate if it was friendly or not. Each participant started with using the letters and intervals, a few participants decided, later on, to just shoot at the mine. According to these

participants, it makes handling the mines easier and faster. This strategy was often the result of participants that struggled to remember the three letters or getting the right interval.

Staying at the starting point

The category staying at the starting point was based upon the code 'not flying'. The commonly incorrect strategy was to stay at the starting point and completely focus on destroying the Space Fortress as many times as possible, instead of flying around the Space Fortress and shoot it while flying. The researcher observed that one participant used this strategy sometimes:

> *''The participant decides to not fly at all and stays at the starting point. He noticed that many points can be achieved in this way.''*

Some participants noticed that the game also gave them points if they were not flying at all, this led to the following question:

> *''I do not see the point in flying around the fortress if it is not shooting back at me, because if I'm staying at the start point I also get points. Why is that?''*

Although some participants noticed that not flying gave them points as well, they still decided to continue flying around the Space Fortress. One participant tried this strategy a few games, but concluded that the game was less exciting this way.

5.4.3 Potential learning problems

While playing the Space Fortress game potential learning problems came to attention. The observations, experiences, and questions provided an overview of the potential learning problems in the Space Fortress game. An example of a potential learning problem was participants that lost their motivation. Table 10 shows the categories of this theme.

Table 10

Categories of potential learning problems

Categories
• Emotions • Losing motivation • Task overload • Impeded progress • Game uncertainty

Emotions

The category emotions was based upon the codes 'frustrated', 'stressed', and 'bored'. The commonly experienced emotions while playing the Space Fortress game were getting bored, stressed, and frustrated. Multiple participants experienced this type of emotions:

> *''I'm getting frustrated because I find this game quite hard.''*
> *''Not great, too many things happened at the same time, I experienced stress because of it and flew into the hyperspace way too often.''*
> *''It becomes more boring for me to play if I'm honest, this can also explain why my scores are getting worse.''*

The participants did not asked specific questions about their emotions.

While observing the participants multiple types of emotions came to attention. Participants often got frustrated when they were not able to handle a specific task, or when they got hit by a mine or the Space Fortress missile. Dying and not knowing how to improve their gameplay also led to many frustrations. Besides frustration, the participants were often stressed as well. The participants were often stressed when they needed to take into account a new task. Adding a new additional task forced the participants to multitask. Getting used to a new task while still performing well on the other tasks was for some participants quite stressful. Furthermore, multiple participants got bored, this happened after playing the full game multiple times. Some participants found this game not motivating enough, while others did not see the added value of the game.

Losing motivation

The category losing motivation was based upon the code 'unmotivated'. Multiple participants experienced a decrease in their motivation at a certain point while playing the Space Fortress game:

> *''I have no motivation left, in my opinion, the number of games I have to play is too much and it takes too long. Besides in my opinion the game is finished because no further tasks are added.''*
> *''The game is lacking an incentive to continue playing, that is not motivating at all.''*
> *''I'm losing my motivation, every game is the same for me, it is not motivating at all.''*

The participants did not asked specific questions about their motivation. However, some questions possibly indicated that they started to lose their motivation:

> *''Can I expect new tasks, or is this the full game?''*
> *''How many games do I still need to play?''*

Multiple participants lost their motivation while playing the game, most of the time this happened in the second training session after the third training part. At this point every game is the same, the participants did not receive any additional or new tasks. While this was one reason why participants lost their motivation, other reasons were the lack of an incentive to continue playing and the number of games that the participants needed to play. Some participants managed to keep motivated, they often challenged themselves to improve their previous scores.

Task overload

The category task overload was based upon the code 'overload'. Multiple participants found the number of games and tasks too much:

> *''I find this game intense because I need to take so many things into account.''*
> *''Playing 8 practice sessions is quite a lot, of course, I will do it, but I could tell that I find it too much.''*
> *''At the beginning, everything went great, but at one point there were too many things that happened at the same time, this influenced my gameplay negatively.''*

As mentioned earlier, the participants asked how many tasks they could expect and how many games they still needed to play, these questions possibly indicated that participants started to lose their motivation, but also that they experienced an overload.

Multiple participants experienced an overload while playing the game. The observations showed that some participants struggled with each new task that was presented to them, the number of tasks that needed to be accomplished at the same time was too much for some participants. While other participants did not directly experience a problem with the multitasking aspect of the game. In addition, participants also experienced an overload due to the number of games people were expected to play. Most participants were not excited to hear that they needed to play 8 practice session games after completing the third training part.

Impeded progress

The category impeded progress was based upon the code 'stuck'. Multiple participants didn't know how to further improve their gameplay:

> *''Not great, I was often hit by the fortress. I am missing feedback from the game itself on how to improve my gameplay. I got the feeling that I make the same mistakes over and over.''*
> *''I don't see that many differences between the games, I'd seems like I got stuck on a point and I do not know how to improve myself to get better.''*
> *''Same as the previous game, it seems like I'm out of my game. I notice that I get more frustrated by this. I think I know what I did wrong but I'm missing concrete tips or feedback on how I can improve or fix this problem. I want insights at the points I struggle with.''*

Participants asked several questions about how they could improve their scores:

> *''How can I improve my scores?''*

The observations showed that people often made the same mistakes, there was a clear need for concrete tips and feedback on how to improve their gameplay. Multiple participants got frustrated and unmotivated because they got stuck in their gameplay and reached the same scores over and over again. Multiple

participants mentioned that they would have liked the idea to have visual examples on how to encounter the several tasks.

Game uncertainty

The category game uncertainty were based upon the codes 'amount of points', and 'when to collect bonus'. Participants often found it unclear how many points could be earned or lost for each task/element in the game, and under which exceptions a bonus could be collected:

> *''I struggle with the fact that I do not have a clue how many points can be earned for each task, and how many points each mistake cost. I focus so much on figuring out these exact numbers that it negatively influences my gameplay.''*

The participants had several questions regarding the game uncertainty:

> *''If you missed the second dollar sign, will the change to collect a bonus be reset? Or will the game continue counting?''*
> *''Can I also collect a bonus when the dollar sign appears three in a row (and I forgot to collect during the second time)?''*
> *''Do you still receive a bonus when you 'die', or is the counter starting on zero again?''*
> *''Are points substracted if you fly counterclockwise?''*
> *''What cost more points, flying into hyperspace or getting hit by a mine?''*
> *''How many points are subtracted when the ship gets hit by a missile or mine?''*
> *''Are points subtracted when I shoot while there are zero missiles left?''*
> *''I find it very unclear what the points exactly say, what is the exact amount of points for each action?''*

The observations showed that people were often confused because several elements were unclear to them. Participants often struggled with finding a good gaming strategy, because they did not know how many points could be earned or lost with each action. The unclarity often led to incorrect strategies.

6. Discussion

Based on the literature discussed in the theoretical framework and introduction, it appeared that feedback is fundamental to the process of gameplaying and is needed for learning processes to take place. When feedback is implemented skills are better ingrained and resilient to skill decay (Kang et al., 2007; Stefanidis et al., 2006). However, the current training element of the SF-AIS is lacking a feedback mechanism. Therefore, a coordinated plan (feedback strategy) under which aspects the feedback should be implemented in the SF-AIS is needed. The main research question in this research was: ''What are the characteristics of a personalized feedback strategy aimed at improving participants' performance in the Space Fortress game?''. This research question is answered by taking into account and categorize the characteristics of the learner and their performances in the Space Fortress game.

6.1 Interpretations and implications

6.1.1 Dropouts

As mentioned before, the Space Fortress training counted multiple dropouts of participants. The performance data of these dropouts showed that participants that quit playing after the first and second training part overall had lower scores than participants that dropped out at the third training part or the practice sessions. This could be an indication that participants who quit playing during the first training sessions struggled more to get familiar with the game and why they decided to quit playing after the first and second training session. However, based on the provided data it is hard to say why people exactly stopped playing. The motivation, self-efficacy, and gaming performance of the participants showed no significant differences.

Further research should consider exploring the reasons why these participants quit playing at an early stage in the game. This could lead to interesting insights and adjusted feedback messages.

6.1.2 Motivation, self-efficacy, and gaming frequency and their relation to participant's performance

Research has shown that self-efficacy and motivation are factors that positively influence learning performances (Narciss, 2013; Yang, Quadir, & Chen, 2016). Therefore, it was expected that self-efficacy and motivation were positively related to participants' performance. In this research positive correlations were found between self-efficacy and game performance and motivation and game performance. However, the correlations in this research showed that overall positive but weak correlations were found between motivation and performance scores and self-efficacy and performance scores. One moderate

positive correlation was found between self-efficacy and the velocity sub-score, but the other correlations between performance scores and these variables were weak. Possible explanations for these weak correlations are the fact that self-efficacy and motivation were only measured once, instead of frequently during the two sessions. Therefore these results might not provide an overall picture of the changes in participants' experienced self-efficacy and motivation. The researcher observed that multiple participants were losing their motivation in the second training session. Towards the end of the second training session, multiple participants said that they started to lose motivation, felt bored, and frustrated. However, the motivation questionnaire was already conducted before the second session, resulting in not taken into account the motivation from the second training session.

The self-efficacy questionnaire was also conducted before the second training session, while the second training session was focused on games that consisted of each task (full games), this was often experienced as more difficult than games that not included each task. Participants possibly had the idea that the given learning tasks in the first session were easy, resulting in the high expectations participants had about their efficacy.

In addition, research had shown that gaming frequency had also a positive influence on performance (Heeter, Lee, Medler, & Magerko, 2011). Therefore, it was also expected that gaming frequency was positively related to participants' performances. In this research, positive strong correlations were found between gaming frequency and participants' performances. Although gaming frequency was only measured once, it is not a variable that was affected by the Space Fortress game. The positive strong correlation between gaming frequency and performances met the expectation of this research.

6.1.3 Learning and performance indicators relevant for the design of personalized feedback

Earlier research provided insights into how verbal protocol data revealed misunderstandings of rules, strategies, the source of erroneous game actions, and goals for improvement (Boot, Sumner, Towne, Rodriguez, & Anders Ericsson, 2017). However, Boot et al. (2017) provided minimal insight into the possible errors, incorrect strategies, and potential learning problems. Therefore, this research conducted a thematic analysis.

After conducting the thematic analysis, three themes could be identified. The themes were based upon multiple categories and codes that emerged from observations, participants' experiences, and questions. The commonly made errors were based upon four categories and nine frequently made mistakes by the participants. This theme provided insight into which mistakes were frequently made

among the participants. The typically made incorrect strategies were based upon four categories and four frequently made mistakes. This theme provided insight into which typical incorrect strategies were used among the participants. The potential learning problem was based upon five categories and eight problems the participants encountered while playing the game. This theme provided insights into the experienced learning problems among the participants.

These outcomes help to identify what the functions and the effect of the feedback message should be. For example, one potential learning problem is losing motivation, one participant mentioned that the game is lacking an incentive to continue playing. According to Narciss (2013), there are three levels of feedback functions, one is on the motivational level, which increases incentive. This potential learning problem can be fixed with feedback that functions on a motivational level.

6.1.4 Group of participants based on performance scores and individual learner characteristics

Research had shown that when the characteristics of the learner are taken into account, tailored feedback messages for a category of learners or an individual learner can be created (Narciss et al., 2014). Earlier research provided insights into the performance differences of high performers and low performers in the Space Fortress game (Lee et al., 2012). However, Lee et al. (2012) only focused on descriptive characteristics like age, proportion male, year of education. Other variables like motivation and gaming frequency were not taken into account. Therefore, this research created different groups of participants, based on the performance scores of each sub-task and participants' gaming frequency, screen time, motivation, the confidence of being able to play the game, if the game was experienced complicated or easy, easy to learn or difficult to learn, and boring or exciting. In this way, tailored feedback messages can be created based on the performances and characteristics of the participants.

The participants were divided into three groups, the high-performing, medium performing, and low-performing group. A clear distinction can be made based on the characteristics and needs of the participants. For example, the low-performing group indicated that they found the game more difficult to learn than the high-performing group and that they often did not understand how to improve their performances. Based upon these needs, performances, and characteristics, different feedback functions, contents, and timing can be given to a specific group. The low-performing group may for example benefit more from receiving detailed feedback. Detailed feedback helps to understand how they need to perform a certain task or how to correct specific errors in their gameplay (Billings, 2012; Shute, 2008).

6.2 Limitations

Although this research found positive results to address the research problem, some limitations should be addressed. First, the sample size in this research was too small to do a regression analysis. Therefore, no clear relationships between the independent variables (motivation, self-efficacy, and game experience) and the dependent variable (gaming performance) could be measured.

Second, the participants were closely observed and asked to evaluate their gameplay and their experiences. Because the researcher closely observed the participants, a possible limitation is that the participants changed their behaviors because of it, which could have led to different outcomes as participants who played the game on their own. This problem is also called reactivity. According to Babbie (2016), ‘’Reactivity is the problem that the subjects of social research may react to the fact of being studied, thus altering their behavior from what it would have been normally’’(p. 292).

Third, participants were asked about their gaming frequency, the question that was used to obtain information about the gaming frequency, was limited to three certain types of games (vector, MOBA, platform). Participants were asked how often they played this type of game in the past 6 months. However, this question does not cover all the types of video games. Some participants mentioned that their gaming frequency was higher than they answered because they played different types of video games. A better solution for measuring gaming frequency was to ask the participants how many hours they spend playing video games, whereas the participant could choose between never, or a few hours per year, month, week, or day, this different measure for gaming frequency can be considered for future research.

Fourth, participants in this research were asked to participate in two training sessions, each session had an average duration of one hour. Although every participant completed both sessions, most participants stated that participating for 2 hours was in their opinion too intense and too long. Multiple participants started to lose their motivation or got distracted in the second session. This could have affected the results. As mentioned earlier, the participants in this research had a custom-built SF-AIS. However, this custom-built SF-AIS needed to be as much as identical as the standard SF-AIS. Therefore, the custom-built SF-AIS contained the same tasks and number of tasks as the standard SF-AIS. Due to this reason, the time of the session was not reduced. Future research should consider reducing the duration time of both sessions or apply a longer waiting time between the two sessions.

Fifth, as mentioned before self-efficacy and motivation were only measured once, instead of frequently during the two sessions. Therefore the results of these variables might not provide an overall picture of the changes in participants' experienced self-efficacy and motivation. Future research should consider measuring participants' self-efficacy and motivation in multiple moments.

Sixth, the three groups that are created in the exploratory data analysis are created based on the 10 participants. As a result, the format of the groups is sample-dependent. Further research should consider using a bigger sample size to create fixed groups.

At last, gender is a demographic characteristic that can relate to learning outcomes. The sample of this study consisted of 8 males and 2 females. Due to the inequality between the number of males and females, it is not possible to consider gender differences in relation to learning outcomes. Although, this was also hard to accomplish due to the small sample size. Future research should consider asking the same amount of males and females to participate in this type of research, to measure the relationship between gender and learning outcomes.

6.3 Recommendations

Based on the results of this research and the characteristics of a feedback strategy as defined by (Narciss, 2012; Narciss et al., 2014), recommendations are made. These recommendations are guidelines for the design and implementation of a personalized feedback strategy that increases participants' performances in the SF-AIS training of NLR.

6.3.1 Function and scope of the feedback

As mentioned in the theoretical framework, the function and scope of feedback can be defined as the purposes or goals the feedback serves. Feedback can have many different functions because it can affect the learning process at various levels (Narciss, 2012; Narciss et al., 2014). The results in this research showed that a distinction can be made between high-performing, middle-performing, and low-performing participants. The participants grouped in each condition also had different needs, problems, and difficulties they faced.

First, the low-performing group often experienced difficulties with their gameplay, errors and incorrect strategies were frequently made. This group had the lowest gaming experiences, which could have led to the fact that this group struggled to understand how to play and improve their gameplay. Therefore this group can benefit from feedback on a cognitive and metacognitive level. Table 11 shows the suggested functions of feedback for the low-performing participants in the Space Fortress game.

Table 11

Suggested feedback functions low-performing group

Feedback function	Definition	Examples
Cognitive level	Feedback on a cognitive level helps to recognize errors, acquire the lacking knowledge, correct incorrect knowledge, and associations, and specify the inaccurate knowledge (Narciss, 2013).	When a participant is repeatedly trying to collect a bonus at the wrong time, awareness of this error can be created by providing feedback on a cognitive level.
Metacognitive level	Feedback on a metacognitive level helps to recognize the incorrect strategies, correct the incorrect strategies, and acquire the missing strategies (Narciss, 2013).	When a participant is shooting at a mine to indicate if it is Foe or Friendly, awareness of this incorrect strategy can be created by providing feedback on a metacognitive level.

Second, the high-performing group often experienced difficulties staying motivated to keep on playing, they missed an incentive to keep on playing. The high-performing group still made mistakes and used incorrect strategies sometimes. However, most of the time the participants in this group knew how to solve their made errors and incorrect strategies. Therefore, this group of participants can particularly benefit from feedback on a motivational level. This type of feedback can help to motivate the participants in this group to keep on playing the game and trying to improve their performances. Table 12 shows the suggested function of feedback for the high-performing participants in the Space Fortress game.

Table 12

Suggested feedback function high-performing group

Feedback function	Definition	Example
Motivational level	This level of feedback helps to increase incentive, effort, and task engagement.	When the performance of a participant is decreasing after playing multiple games in a row, feedback on a motivational level can be provided to increase the effort and task engagement.

Third, the participants of the middle-performing group scored below average on 50 percent of sub-tasks, while on the other 50 percent they scored above average. To improve the performances on the tasks the participants in this group struggle with, feedback on the cognitive level and metacognitive level is advisable. To improve the performances on the tasks where the participants scored above average, feedback on the motivational level is recommended. Depending on the performance score of each learning task, different types of feedback functions can be functional.

In short, the first recommendation is to determine the function of the feedback based on the different groups. Tailored feedback messages can be created and given based on the needs and performances of the different groups.

6.3.2 Presentation and timing of the feedback

A large variety exists in the way feedback is communicated or presented. To present the feedback content multiple aspects like the feedback timing (e.g. delayed or immediate) need to be taken into account (Narciss, 2013; Shute, 2008). Based upon the level of achievement of the participants the most suitable feedback timing can be determined.

Although a clear distinction could be made between high, middle, and low-performing participants, at the start of the SF-AIS training, every participant was unknown with the Space Fortress game and needed to learn several new tasks. When a new task was introduced, each participant struggled at the beginning to control this new task. Some participants were able to control this new task faster than others, but their start was often the same. Johnson et al. (2017) stated that when novices learn new procedures in serious games, the use of immediate feedback should be considered. Therefore, it is

recommended to start with immediate feedback for each participant in the training parts where a new task is introduced.

After each task is introduced, the timing can be based on participants' capabilities. According to Shute (2008), it is better to use immediate feedback when a task is perceived as difficult, this can function as a safety net for the learner and avoids possible frustrations. To prevent possible annoyance and feedback intrusion, delayed feedback is better for tasks which are perceived simple by the learner (Shute, 2008). However, the difficulty of a task is relative to the learners' capabilities. Therefore, the second recommendation regarding the timing of the feedback is to consider using delayed feedback when participants score above average, and to use immediate feedback when participants score below average.

6.3.3 Feedback content

Billings (2012); Johnson et al. (2017); Serge et al. (2013) stated that performances improve when novice learners are provided with detailed feedback. However, when the mastery of the player increases, the player could get distracted by feedback, if it provides information that they already know. When a player's skills increases the level of detailed feedback should decrease (Serge et al., 2013).

Therefore, the first recommendation regarding the feedback content is to provide the participants in the low-performing group with detailed and specific feedback on how to perform a certain task, or how to correct specific errors in their gameplay. The participants receive principle-based feedback, where the quality of their performance is described. This type of feedback is also called process feedback (i.e., explanatory feedback (Johnson et al., 2017)). According to Johnson et al. (2017), novices learn more deeply if they receive explanatory feedback. Table 13 shows the suggested types of explanatory feedback and examples for the low-performing participants in the Space Fortress game.

Table 13

Suggested explanatory feedback types

Feedback type	Definition	Examples
Informational	Informational feedback is focused on providing some information to improve the general understanding or how to obtain the correct answer	''Don't forget to collect a bonus, pay attention to the $ signs''
Topic Specific	Topic Specific feedback is focused on providing specific information about a topic, it leads the learner through the correct action or answer	''You can collect a bonus when the $ sign appeared twice in a row, press K for munition and L for free points''
Response Specific	Response feedback addresses why a certain strategy or action is correct or incorrect.	''You tried to collect a bonus after the $ sign appeared one time, you can only collect a bonus when the $ sign appeared twice in a row''

Note. Table adapted from Johnson et al. (2017). The same notation of feedback types and definitions is used. Different examples are used (examples are linked to the Space Fortress content).

The second recommendation regarding the feedback content is to provide the participants in the high-performing with feedback that challenges them. Therefore, it is recommended to provide outcome feedback and more general feedback, such as prompts, cues, and hints to improve their performances. Table 14 shows the suggested types of outcome and explanatory feedback and examples for the high-performing participants in the Space Fortress game.

Table 14

Suggested outcome and explanatory feedback types

Feedback type		Definition	Example
Outcome	Velocity	Participants performance is compared to their performances in the previous game	''Your performance on the SPEED sub-score declined compared to the previous game''
	Normative	The ranking of the participant is compared to the other participants	''Your performance is better than 90% of the other participants''
Explanatory	Hints/Prompts/Cues	Hints/prompts/cues are provided to guide the participants to the right strategy/approach, without explicitly saying what they need to do	''Here is a hint: To identify a mine, you should take a look between the VLCTY and INTRVL score''

Note. Table adapted from Johnson et al. (2017). The same notation of feedback types and definitions is used. Different examples are used (examples are linked to the Space Fortress content).

To implement normative outcome feedback for the high-performing participants, a leaderboard for each sub-task of the Space Fortress game can be considered. With the use of a leaderboard, participants can compare their performances against that of other participants (Sun, Jones, Traca, & Bos, 2015). A leaderboard presents the score of the participants by an ordered ranking. When a participant is placed high on the leaderboard, the game affirms that the participant is competent to play a certain task compared to the other participants. This should lead to a greater feeling of competence. When the feeling of competence increases, the game enjoyment should increase (Velez, Ewoldsen, Hanus, Song, & Villarreal, 2018). Leaderboard can create an external motivation, participants experiences of the gameplay shift from an internal desire to an external force (competition driven by other participants) (Velez et al., 2018). During

the observations in this research, it became clear that the participants in the high-performing group lacked internal desire to play the game, a leaderboard can help these participants to keep motivated and improve their performances. However, when a participant got a low leaderboard position, it can negatively affect competence and decrease participants' enjoyment (Velez et al., 2018). Therefore, the leaderboard should be optional, the participants need to have the choice if they want to see the leaderboard or not.

The participants in the middle-performing group may benefit from more detailed feedback on the task they struggle with, such as described for the low-performing group. For the task where they scored above average, the same approach as for the high-performing group is recommended.

The last recommendation is to design the feedback based upon the results of the thematic analysis in this research, which showed the frequently made errors, incorrect strategies, and potential learning problems.

7. Conclusion

To improve participant's performances and prevent possible decay or loss of skills due to the lack of feedback, the adaptation of personalized feedback is needed.

This research aimed to investigate what characteristics of a personalized feedback strategy could increase participants' performance in the Space Fortress game. Several studies created frameworks and guidelines for feedback strategies (Narciss, 2013; Narciss & Huth, 2004; Shute, 2008). These frameworks identified important characteristics of a personalized feedback strategy, upon which this study focused. By identifying the individual learner characteristics, the learning and performance indicators, and homogeneous groups relevant for the design of personalized feedback. As stated in literature the presentation, content, and function format of the feedback message should be driven by considerations of the learner characteristics and instructional goals, to maximize the informative value of the feedback (Narciss & Huth, 2004; Shute, 2008).

Based on a qualitative and quantitative analysis of participants' learner characteristics, performances, and experiences, it can be concluded that different learning problems, errors, incorrect strategies, and categories of participants exist upon which tailored personalized feedback can be created. The results indicate that the different categories of participants encounter different errors, incorrect strategies, and learning problems regarding playing the game. To improve participants' performances, recommendations are made regarding the feedback function, presentation, and content.

Further research is needed to implement the recommended feedback strategy and to determine the effects of this personalized feedback on participants' performances in the SF-AIS of NLR. Besides, this research was mainly conducted from the perspective of the highest score achieved. A learning tool can also be viewed from the perspective of the progress of a participant. A bad gamer can for example make more progress than an experienced gamer. Further research should consider taking into account the progress of a participant while designing personalized feedback.

Reference list

Arthur Jr, W., Bennett Jr, W., Stanush, P. L., & McNelly, T. L. (1998). Factors That Influence Skill Decay and Retention: A Quantitative Review and Analysis. *Human Performance, 11*(1), 57-101. doi:10.1207/s15327043hup1101_3

Babbie, E. R. (2016). *The practice of social research*.

Billings, D. R. (2012). Efficacy of Adaptive Feedback Strategies in Simulation-Based Training. *Military Psychology, 24*(2), 114-133. doi:10.1080/08995605.2012.672905

Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation? *Qualitative Health Research, 26*. doi:10.1177/1049732316654870

Bolton, A. (2006). Immediate Versus Delayed Feedback In Simulation Based Training: Matching Feedback Delivery Timing To The Cognitive Demands Of Th.

Boot, W. R., Sumner, A., Towne, T. J., Rodriguez, P., & Anders Ericsson, K. (2017). Applying Aspects of the Expert Performance Approach to Better Understand the Structure of Skill and Mechanisms of Skill Acquisition in Video Games. *Topics in Cognitive Science, 9*(2), 413-436. doi:https://doi.org/10.1111/tops.12230

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa

Burgers, C., Eden, A., van Engelenburg, M. D., & Buningh, S. (2015). How feedback boosts motivation and play in a brain-training game. *Computers in Human Behavior, 48*, 94-103. doi:https://doi.org/10.1016/j.chb.2015.01.038

Cameron, B., & Dwyer, F. (2005). The Effect of Online Gaming, Cognition and Feedback Type in Facilitating Delayed Achievement of Different Learning Objectives. *Journal of Interactive Learning Research, 16*(3), 243-258.

Charles, D., Charles, T., McNeill, M., Bustard, D., & Black, M. (2011). Game-based feedback for educational multi-user virtual environments. *British Journal of Educational Technology, 42*(4), 638-654. doi:https://doi.org/10.1111/j.1467-8535.2010.01068.x

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a New General Self-Efficacy Scale. *Organizational Research Methods, 4*(1), 62-83. doi:10.1177/109442810141004

Clariana, R. B., Wagner, D., & Murphy, L. C. R. (2000). Applying a connectionist description of feedback timing. *Educational Technology Research and Development, 48*(3), 5-22.

Corbalan, G., Kester, L., & van Merriënboer, J. (2009). Dynamic task selection: Effects of feedback and learner control on efficiency and motivation. *Learning and Instruction, 19*(6), 455-465. doi:https://doi.org/10.1016/j.learninstruc.2008.07.002

Drachsler, H., & Kirschner, P. A. (2012). Learner Characteristics. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 1743-1745). Boston, MA: Springer US.

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education, 67*, 156-167. doi:https://doi.org/10.1016/j.compedu.2013.02.019

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*: Thomson Brooks/Cole Publishing Co.

Evans, M. A., Pruett, J., Chang, M., & Nino, M. (2014). Designing Personalized Learning Products for Middle School Mathematics: The Case for Networked Learning Games. *Journal of Educational Technology Systems, 42*(3), 235-254. doi:10.2190/ET.42.3.d

Goodman, J. S., Wood, R. E., & Chen, Z. (2011). Feedback specificity, information processing, and transfer of training. *Organizational Behavior and Human Decision Processes, 115*(2), 253-267. doi:https://doi.org/10.1016/j.obhdp.2011.01.001

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion, 24*(3), 175-213. doi:10.1023/A:1005614228250

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research, 77*(1), 81-112. doi:10.3102/003465430298487

Heeter, C., Lee, Y.-H., Medler, B., & Magerko, B. (2011). *Beyond player types: gaming achievement goal*. Paper presented at the Proceedings of the 2011 ACM SIGGRAPH Symposium on Video Games, Vancouver, British Columbia, Canada. https://doi.org/10.1145/2018556.2018565

Johnson, C. I., Bailey, S. K. T., & Van Buskirk, W. L. (2017). Designing Effective Feedback Messages in Serious Games and Simulations: A Research Review. In P. Wouters & H. van Oostendorp (Eds.), *Instructional Techniques to Facilitate Learning and Motivation of Serious Games* (pp. 119-140). Cham: Springer International Publishing.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*(4-5), 528-558. doi:10.1080/09541440601056620

Killingsworth, S. S., Clark, D. B., & Adams, D. M. (2015). Self-Explanation and Explanatory Feedback in Games: Individual Differences, Gameplay, and Learning. *International Journal of Education in Mathematics, Science and Technology, 3(3)*, 162-186.

Kim, J. W., Ritter, F. E., & Koubek, R. J. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science, 14*(1), 22-37. doi:10.1080/1464536X.2011.573008

Kluge, A., & Frank, B. (2014). Counteracting skill decay: four refresher interventions and their effect on skill and knowledge retention in a simulated process control task. *Ergonomics, 57*(2), 175-190. doi:10.1080/00140139.2013.869357

Lee, H., Boot, W. R., Basak, C., Voss, M. W., Prakash, R. S., Neider, M., . . . Kramer, A. F. (2012). Performance gains from directed training do not transfer to untrained tasks. *Acta psychologica, 139*(1), 146-158. doi:https://doi.org/10.1016/j.actpsy.2011.11.003

Lieffijn, D. (2020). *Modelling skill retention in Space Fortress using machine learning.* (Bachelor thesis), Utrecht University Retrieved from http://dspace.library.uu.nl/bitstream/handle/1874/396279/Bachelor%20thesis%20AI%20%5BLieffijn%2C%205587301%5D.pdf?sequence=1&isAllowed=y

Lookadoo, K. L., Bostwick, E. N., Ralston, R., Elizondo, F. J., Wilson, S., Shaw, T. J., & Jensen, M. L. (2017). "I Forgot I Wasn't Saving the World": the Use of Formative and Summative Assessment in Instructional Video Games for Undergraduate Biology. *Journal of Science Education and Technology, 26*(6), 597-612. doi:10.1007/s10956-017-9701-5

Mané, A., & Donchin, E. (1989). The space fortress game. *Acta psychologica, 71*(1-3), 17-22.

Mason, B., & Bruning, R. (2001). Providing Feedback in Computer-based Instruction: What the Research Tells Us. *Center for Instructional Innovation, 15*.

Mayer, R. E., & Johnson, C. I. (2010). Adding Instructional Features That Promote Learning in a Game-Like Environment. *Journal of Educational Computing Research, 42*(3), 241-265. doi:10.2190/EC.42.3.a

Narciss, S. (2012). Feedback Strategies. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 1289-1293). Boston, MA: Springer US.

Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning environments on the basis of the Interactive Tutoring Feedback Model. *Digital Education Review*(23), 7-26.

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. *Instructional design for multimedia learning, 181195*.

Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichelmann, A., Goguadze, G., & Melis, E. (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education, 71*, 56-76. doi:https://doi.org/10.1016/j.compedu.2013.09.011

Ricci, K. E., Salas, E., & Cannon-Bowers, J. A. (1996). Do Computer-Based Games Facilitate Knowledge Acquisition and Retention? *Military Psychology, 8*(4), 295-307. doi:10.1207/s15327876mp0804_3

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2020). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior, 107*, 105512. doi:https://doi.org/10.1016/j.chb.2018.05.004

Serge, S. R., Priest, H. A., Durlach, P. J., & Johnson, C. I. (2013). The effects of static and adaptive performance feedback in game-based training. *Computers in Human Behavior, 29*(3), 1150-1158. doi:https://doi.org/10.1016/j.chb.2012.10.007

Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research, 78*(1), 153-189. doi:10.3102/0034654307313795

Smits, M. H. S. B., Boon, J., Sluijsmans, D. M. A., & van Gog, T. (2008). Content and timing of feedback in a web-based learning environment: effects on learning as a function of prior knowledge. *Interactive Learning Environments, 16*(2), 183-193. doi:10.1080/10494820701365952

Stefanidis, D., Korndorffer, J. R., Markley, S., Sierra, R., & Scott, D. J. (2006). Proficiency Maintenance: Impact of Ongoing Simulator Training on Laparoscopic Skill Retention. *Journal of the American College of Surgeons, 202*(4), 599-603. doi:https://doi.org/10.1016/j.jamcollsurg.2005.12.018

Stern, Y., Blumen, H. M., Rich, L. W., Richards, A., Herzberg, G., & Gopher, D. (2011). Space Fortress game training and executive control in older adults: A pilot intervention. *Aging, Neuropsychology, and Cognition, 18*(6), 653-677. doi:10.1080/13825585.2011.613450

Stobbeleir, K. E. M. d., Ashford, S. J., & Buyens, D. (2011). Self-Regulation of Creativity at Work: The Role of Feedback-Seeking Behavior in Creative Performance. *Academy of Management Journal, 54*(4), 811-831. doi:10.5465/amj.2011.64870144

Sun, E., Jones, B., Traca, S., & Bos, M. W. (2015). *Leaderboard Position Psychology: Counterfactual Thinking*. Paper presented at the Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, Seoul, Republic of Korea. https://doi.org/10.1145/2702613.2732732

Van der Pal, J., & Toubman, A. (2020). *An Adaptive Instructional System for the Retention of Complex Skills*, Cham.

Van Mourik, D. (2020). *The effects of scaffolding and feedback adaptive to the characteristics of the learner on learning progress and performance for the purpose of personalized learning.* (Master's thesis), Open Universiteit Retrieved from https://research.ou.nl/en/studentTheses/the-effects-of-scaffolding-and-feedback-adaptive-to-the-character

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior, 27*(1), 118-130. doi:https://doi.org/10.1016/j.chb.2010.07.038

Velez, J. A., Ewoldsen, D. R., Hanus, M. D., Song, H., & Villarreal, J. A. (2018). Social Comparisons and Need Fulfillment: Interpreting Video Game Enjoyment in the Context of Leaderboards. *Communication Research Reports, 35*(5), 424-433. doi:10.1080/08824096.2018.1525352

Vlasblom, J. I. D., Pennings, H. J. M., van der Pal, J., & Oprins, E. A. P. B. (2020). Competence retention in safety-critical professions: A systematic literature review. *Educational Research Review, 30*, 100330. doi:https://doi.org/10.1016/j.edurev.2020.100330

Weaver, S. J., Newman-Toker, D. E., & Rosen, M. A. (2012). Reducing cognitive skill decay and diagnostic error: Theory-based practices for continuing education in health care. *Journal of Continuing Education in the Health Professions, 32*(4), 269-278. doi:https://doi.org/10.1002/chp.21155

Yang, J. C., Quadir, B., & Chen, N.-S. (2016). Effects of the Badge Mechanism on Self-Efficacy and Learning Performance in a Game-Based English Learning Environment. *Journal of Educational Computing Research, 54*(3), 371-394. doi:10.1177/0735633115620433

Appendices

Appendix A: Consent form

Please tick the appropriate boxes	Yes	No
Taking part in the study		
I have read and understood the study information dated / / , or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	○	○
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	○	○
I understand that taking part in the study involves an audio-recorded interview, questionnaires, playing the Space Fortress game and doing the Space Fortress rules quiz. The audio recording will be transcribed as text, the recording will be destroyed after finishing the thesis.	○	○
Use of the information in the study		
I understand that information I provide will be used for publications and reports	○	○
I understand that personal information collected about me that can identify me, such as [e.g. my name or where I live], will not be shared beyond the study team.	○	○
I agree to be audio recorded	○	○
Future use and reuse of the information by others		
I give permission for the data that I provide to be archived on an active NLR server so it can be used for future research and learning. The game data will be stored separately from your personal information, such as your name, email address. Personal data we gather will only be available to NLR employees involved in the project. The data will be stored on an active NLR server until the end of the project (December 2021) or its successor project (December 2025). After this period, the data will be stored for a period of 5 years on a tape, after which all recorded data will be destroyed.	○	○
I agree that my information may be shared with other researchers for future research studies that may be similar to this study or may be completely different. The information shared with other researchers will not include any information that can directly identify me. Researchers will not contact me for additional permission to use this information.	○	○

I give the researchers permission to keep my contact information and to contact me for future research projects. ○ ○

Signatures

_____	_____	_____
Name of participant	Signature	Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

_____	_____	_____
Researcher name	Signature	Date

Study contact details for further information:

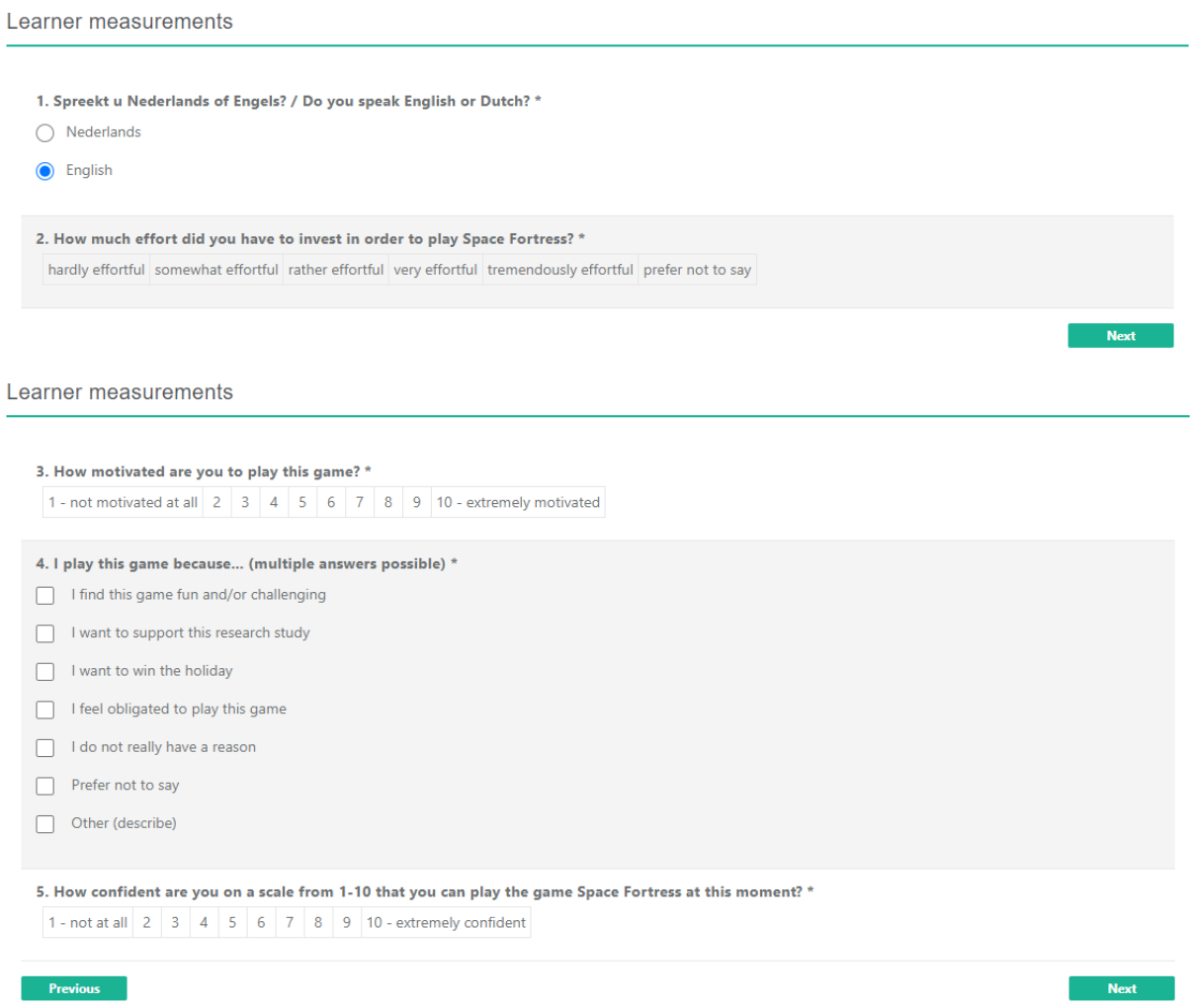
Iris Kroos

iriskroos@gmail.com

Contact Information for Questions about Your Rights as a Research Participant

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-bms@utwente.nl

Appendix B: Learner measurement questionnaire



Learner measurements

1. Spreekt u Nederlands of Engels? / Do you speak English or Dutch? *

- ○ Nederlands
- ◉ English

2. How much effort did you have to invest in order to play Space Fortress? *

hardly effortful | somewhat effortful | rather effortful | very effortful | tremendously effortful | prefer not to say

Next

Learner measurements

3. How motivated are you to play this game? *

1 - not motivated at all | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 - extremely motivated

4. I play this game because... (multiple answers possible) *

- ☐ I find this game fun and/or challenging
- ☐ I want to support this research study
- ☐ I want to win the holiday
- ☐ I feel obligated to play this game
- ☐ I do not really have a reason
- ☐ Prefer not to say
- ☐ Other (describe)

5. How confident are you on a scale from 1-10 that you can play the game Space Fortress at this moment? *

1 - not at all | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 - extremely confident

Previous

Next

6. How do you feel today? *

Extremely alert	Very alert	Alert	Rather alert	Neither alert nor sleepy	Some signs of sleepiness	Sleepy, but no difficulty remaining awake	Sleepy, some effort to keep alert	Extremely sleepy, fighting sleep

7. How do you feel today? *

- ◯ sad, tired
- ◯ neutral, content
- ◯ engaged, excited
- ◯ frustrated, stressed
- ◯ prefer not to say

8. How do you feel when playing this game? *

- ◯ sad, tired
- ◯ neutral, content
- ◯ engaged, excited
- ◯ frustrated, stressed
- ◯ prefer not to say

9. How do you feel today? *

1 - very ill	2	3	4	5	6	7 - very healthy	prefer not to say

10. Are you in a distraction free (1) or distraction rich (10) environment? Think about: distractions, noise, other people in the room vs quiet, clean desk, etc. *

1 - distraction free	2	3	4	5	6	7	8	9	10 - distraction rich

2. What is your date of birth? *

3. What is your gender? *

- ◯ male
- ◯ female
- ◯ other
- ◯ prefer not to say

4. In which country do you live? *

- ◯ The Netherlands
- ◯ Prefer not to say
- ◯ Other:

5. What is your highest level of education? *

- ◯ primary school
- ◯ high school
- ◯ college
- ◯ university
- ◯ prefer not to say

6. How many hours a day do you spend behind a screen? *

- ◯ less than 2 hours
- ◯ 2 – 4 hours
- ◯ 4 – 6 hours
- ◯ more than 6 hours

- **Platformer games** are games in which the player controls a character which has to jump from platform to platform, and can encounter enemies and points. One famous platformer game is Super Mario Bros.
- **Vector arcade games** contain minimal graphics and have been played especially in the '80. One famous vector arcade game is Pac-Man.
- **Moba games** are about a war between two teams. The player controls one of these teams in a strategic manner. One famous moba game is League of Legends.

7. How often have you played these games over the past half year? *

	never	1x a month or less	1-4x a month	>1x a week	prefer not to say
Platformer games	◯	◯	◯	◯	◯
Vector arcade games	◯	◯	◯	◯	◯
Moba games	◯	◯	◯	◯	◯

Vorige Volgende

8. What is your profession? *

9. At work I perform the following skills: *

	never	1x a month or less	monthly	weekly	daily	prefer not to say
Operating vehicles, instruments or machines	◯	◯	◯	◯	◯	◯
Standard desktop activities, such as emailing, writing, programming, drawing, use of internet	◯	◯	◯	◯	◯	◯
Monitoring processes (requiring little input from my side)	◯	◯	◯	◯	◯	◯
Controlling processes (requiring constant input from my side)	◯	◯	◯	◯	◯	◯
Making decisions	◯	◯	◯	◯	◯	◯
Working with procedures	◯	◯	◯	◯	◯	◯

10. Do you have a military background? *

◯ yes

◯ no

◯ prefer not to say

11. Do you receive refresher trainings at work? *

◯ yes

◯ no

◯ prefer not to say

12. In your daily tasks, is your response usually required within seconds?

◯ yes

◯ no

2. Indicate to what extent you agree with the following statements: *

	1 (not at all)	2	3	4 (a bit)	5	6	7 (completely
The goal and the rules of the game were clear to me	◯	◯	◯	◯	◯	◯	◯
The instructions were easy to understand	◯	◯	◯	◯	◯	◯	◯
The instructions were sufficient	◯	◯	◯	◯	◯	◯	◯
I received enough feedback about my gameplay	◯	◯	◯	◯	◯	◯	◯
I have mastered the game up to a good level	◯	◯	◯	◯	◯	◯	◯
I am satisfied about my game results	◯	◯	◯	◯	◯	◯	◯
I would want to play the game for a longer period	◯	◯	◯	◯	◯	◯	◯
I want to keep on participating in this research study	◯	◯	◯	◯	◯	◯	◯

3. How would you rate this Space Fortress training? *

	1	2	3	4	5	6	7
annoying (1) --- enjoyable (7)	○	○	○	○	○	○	○
not understandable (1) --- understandable (7)	○	○	○	○	○	○	○
easy to learn (1) --- difficult to learn (7)	○	○	○	○	○	○	○
valuable (1) --- inferior (7)	○	○	○	○	○	○	○
boring (1) --- exciting (7)	○	○	○	○	○	○	○
not interesting (1) --- interesting (7)	○	○	○	○	○	○	○
fast (1) --- slow (7)	○	○	○	○	○	○	○
good (1) --- bad (7)	○	○	○	○	○	○	○
complicated (1) --- easy (7)	○	○	○	○	○	○	○
unlikable (1) --- pleasing (7)	○	○	○	○	○	○	○
unpleasant (1) --- pleasant (7)	○	○	○	○	○	○	○
motivating (1) --- demotivating (7)	○	○	○	○	○	○	○
inefficient (1) --- efficient (7)	○	○	○	○	○	○	○
clear (1) --- confusing (7)	○	○	○	○	○	○	○
impractical (1) --- practical (7)	○	○	○	○	○	○	○
organized (1) --- cluttered (7)	○	○	○	○	○	○	○
attractive (1) --- unattractive (7)	○	○	○	○	○	○	○
friendly (1) --- unfriendly (7)	○	○	○	○	○	○	○

4. How would you rate this Space Fortress training in total? *

1	2	3	4	5	6	7	8	9	10	prefer not to say

Appendix C: Situational Motivation Scale

Fill in these questions while thinking about the game.

13. I do this activity because... *

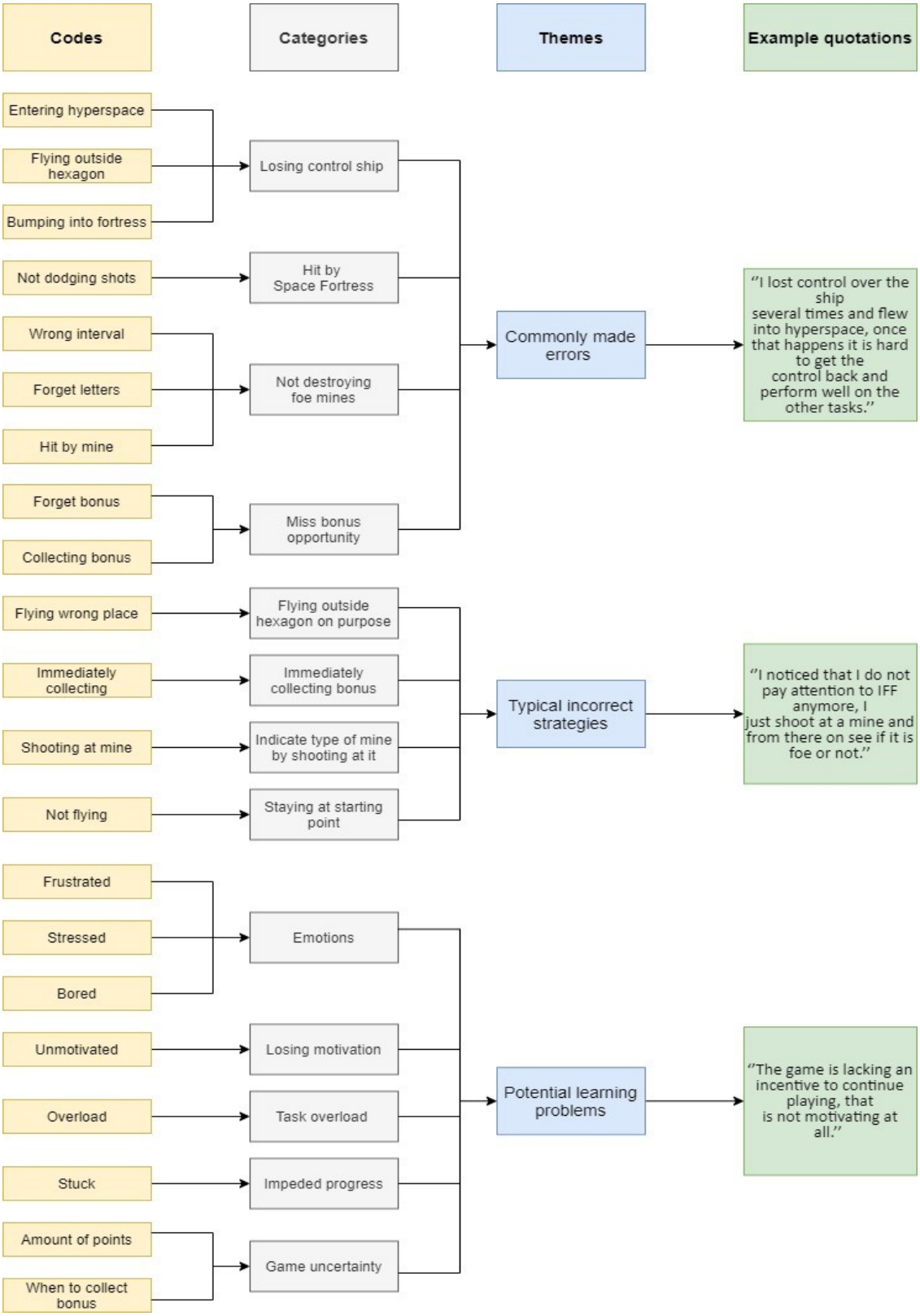
	1 - not at all true	2	3	4	5	6	7 - very true
Because I think that this activity is interesting	○	○	○	○	○	○	○
Because I think that this activity is pleasant	○	○	○	○	○	○	○
Because this activity is fun	○	○	○	○	○	○	○
Because I feel good when doing this activity	○	○	○	○	○	○	○
Because I am doing it for my own good	○	○	○	○	○	○	○
Because I think that this activity is good for me	○	○	○	○	○	○	○
By personal decision	○	○	○	○	○	○	○
Because I believe that this activity is important for me	○	○	○	○	○	○	○
Because I am supposed to do it	○	○	○	○	○	○	○
Because it is something that I have to do	○	○	○	○	○	○	○
Because I don't have any choice	○	○	○	○	○	○	○
Because I feel that I have to do it	○	○	○	○	○	○	○
There may be good reasons to do this activity, but personally I don't see any	○	○	○	○	○	○	○
I do this activity but I am not sure if it is worth it	○	○	○	○	○	○	○
I don't know; I don't see what this activity brings me	○	○	○	○	○	○	○
I do this activity, but I am not sure it is a good thing to pursue it	○	○	○	○	○	○	○

Appendix D: New General Self-Efficacy Scale

16. Fill in these questions while thinking about the game. *

	1 - completely disagree	2	3	4	5 - completely agree	prefer not to say
I will be able to achieve most of the goals that I have set for myself.	◯	◯	◯	◯	◯	◯
When facing difficult tasks, I am certain that I will accomplish them.	◯	◯	◯	◯	◯	◯
In general, I think that I can obtain outcomes that are important to me.	◯	◯	◯	◯	◯	◯
I believe I can succeed at most any endeavor to which I set my mind	◯	◯	◯	◯	◯	◯
I will be able to successfully overcome many challenges	◯	◯	◯	◯	◯	◯
I am confident that I can perform effectively on many different tasks.	◯	◯	◯	◯	◯	◯
Compared to other people, I can do most tasks very well	◯	◯	◯	◯	◯	◯
Even when things are tough, I can perform quite well.	◯	◯	◯	◯	◯	◯

Appendix E: Thematic analysis



Appendix F: Overview used codes

E: Experience participant
O: Observation researcher
Q: Question participant

Code	Definition	Example
Entering hyperspace	When the spaceship flies into hyperspace.	**E:** ''I lost control over the ship several times and flew into hyperspace, once that happens it is hard to get the control back and perform well on the other tasks.''
Flying outside hexagon	When the spaceship flies outside the hexagon.	**E:** ''several times I flew outside the hexagon area''
Bumping into fortress	When the spaceship bumps into the space fortress.	**E:** ''Sometimes I bumped into the fortress.'' **E:** ''I notice knowing the fortress would not shoot back that I bump into the fortress more often.''
Not dodging shots	When the participant was not able to dodge the shots of the space fortress with the spaceship	**E:** ''I got hit several times by the missiles of the fortress.'' **E:** ''I struggled with dodging the shots of the fortress, almost every shot hit me.''
Wrong interval	When the participant is not able to get the right interval time for the J key	**E:** ''I troubled to handle the foe mines, my average interval for the J key was 128, this is way too fast.''
Forget letters	When the participant forget or forget to remember the three letters indicating a foe mine	**E:** ''[…] I also forgot the three letters indicating a foe mine'' **E:** ''I forgot to remember the three letters indicating a foe mine.''
Hit by mine	When the spaceship is hit by a mine	**E:** ''I got hit several times by the mines.'' **E:** ''Because of the mines, I died several times.''

Forget bonus	When the participant forgets to collect a bonus	**E:** ''Sometimes I forget to collect a bonus.''
Collecting bonus	When the participant does not pay attention to collect a bonus	**E:** ''I noticed that I only focused on collecting the bonus at the beginning of the game, after that, I did not pay attention to this part of the game anymore, there were still a lot of other tasks I needed to focus on.''
Flying wrong place	When the participant flies the spaceship outside the hexagon on purpose, to avoid being shot	**E:** ''I notice that I'm flying outside the hexagon and the square, to easily dodge the shots from the fortress.''
Immediately collecting	When the participant tries to collect a bonus instantly, when a dollar sign appears on the screen, not knowing how many times it occurred	**E:** ''Once I see a dollar sign I immediately try to collect a bonus, but most of the time that does not work.''
Shooting at mine	When the participant shoots at a mine to indicate if it is a foe mine or friendly mine, instead of remembering the three letters indicating a foe mine.	**E:** ''I noticed that I do not pay attention to IFF anymore, I just shoot at a mine and from there on see if it is foe or not.''
Not flying	When the participant decides not to fly at all, but instead is staying at the starting point and is completely focused on destroying the fortress as many times as possible.	**O:** ''The participant decides to not fly at all and stays at the starting point. He noticed that many points can be achieved in this way.''
Frustrated	When the participant gets frustrated because of playing the game	**E:** ''I'm getting frustrated because I find this game quite hard.'' **E: ''**It seems like I'm out of my game. I notice that I get more frustrated by this. I think I know what I did wrong but I'm missing concrete tips or feedback on how I can improve or fix this problem. I want insights at the points I struggle with.''

		E: ‘’I struggle with handling the foe mines, this makes me frustrated.’’
Stressed	When the participant is getting stressed while playing the game	**E:** ‘’The bonus task makes playing the game a lot harder, there many things that I need to take into account. This makes me stressed. For me, it was more helpful if the instructions provided me with an overview and explanation where I could find everything’’ **E:** ‘’Not great, too many things happened at the same time, I experienced stress because of it and flew into the hyperspace way too often.’’
Bored	When the participant is getting bored	**E:** ‘’It becomes more boring for me to play if I’m honest, this can also explain why my scores are getting worse.’’ **E:** ‘’I don’t see what the purpose and the added value of this game should be, the last few sessions were quite boring and not motivating for me.’’
Unmotivated	When the participant is losing motivation	**E:** ‘’I have no motivation left, in my opinion, the number of games I have to play is too much and it takes too long. Besides in my opinion the game is finished because no further tasks are added.’’ **E:** ‘’The game is lacking an incentive to continue playing, that is not motivating at all.’’ **E:** ‘’I’m losing my motivation, every game is the same for me, it is not motivating at all.’’
Overload	When the amount of games and exercises is too much for the participant	**E:** ‘’Playing 8 practice sessions is quite a lot, of course, I will do it, but I could tell that I find it too much.’’

Stuck	When the participant does not know how to further improve his or her gameplay	**E:** ''At the beginning, everything went great, but at one point there were too many things that happened at the same time, this influenced my gameplay negatively.'' **E:** ''I find this game intense because I need to take so many things into account.'' **E:** I don't see that many differences between the games, I'd seems like I got stuck on a point and I do not know how to improve myself to get better.'' **E:** ''Not great, I was often hit by the fortress. I am missing feedback from the game itself on how to improve my gameplay. I got the feeling that I make the same mistakes over and over.''
Amount of points	When the participant finds it unclear how many points can be earned or lost for each task/element of the game	**E:** ''I find it very unclear what the points exactly say, what is the exact amount of points for each action?'' **E:** ''I struggle with the fact that I do not have a clue how many points can be earned for each task, and how many points each mistake cost. I focus so much on figuring out these exact numbers that it negatively influences my gameplay.''
When to collect bonus	When the participant finds it unclear under which exceptions a bonus can be collected.	**Q:** ''Do you still receive a bonus when you 'die', or is the counter starting on zero again?'' **Q:** ''If you missed the second dollar sign, will the change to collect a bonus be reset? Or will the game continue counting?'' **Q:** ''Can I also collect a bonus when the dollar sign appears three in a row (and I forgot to collect during the second time)?''

Appendix G: Graphs exploratory data analysis

Figure 5

Average score and mean scores of the control sub-score

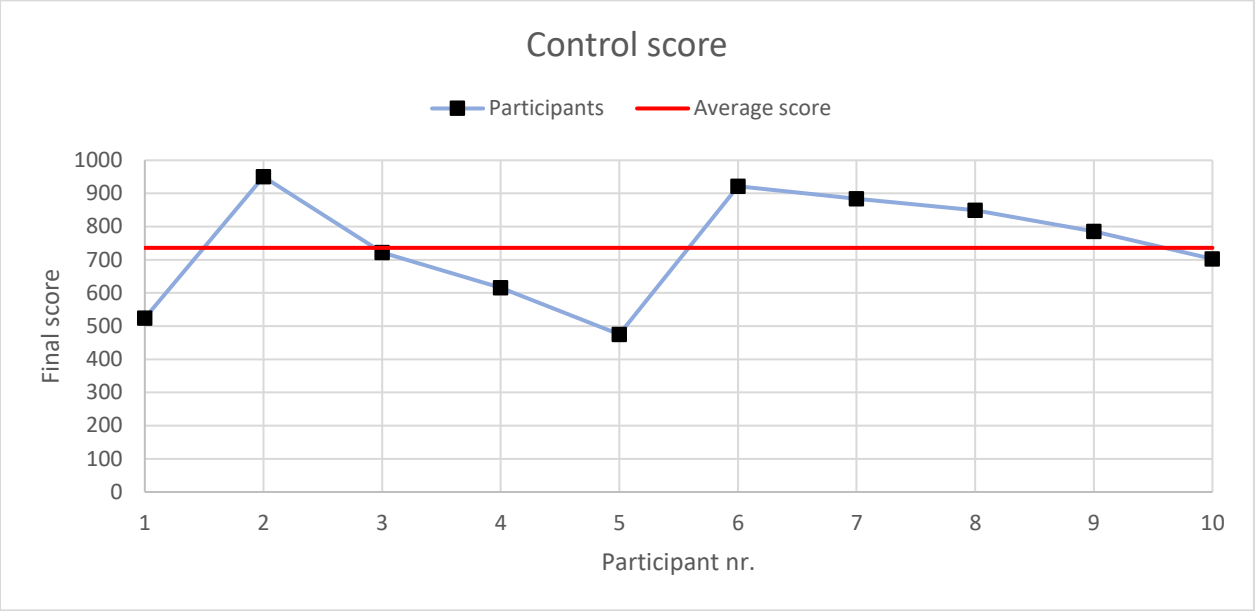


Figure 6

Average score and mean scores of the points sub-score

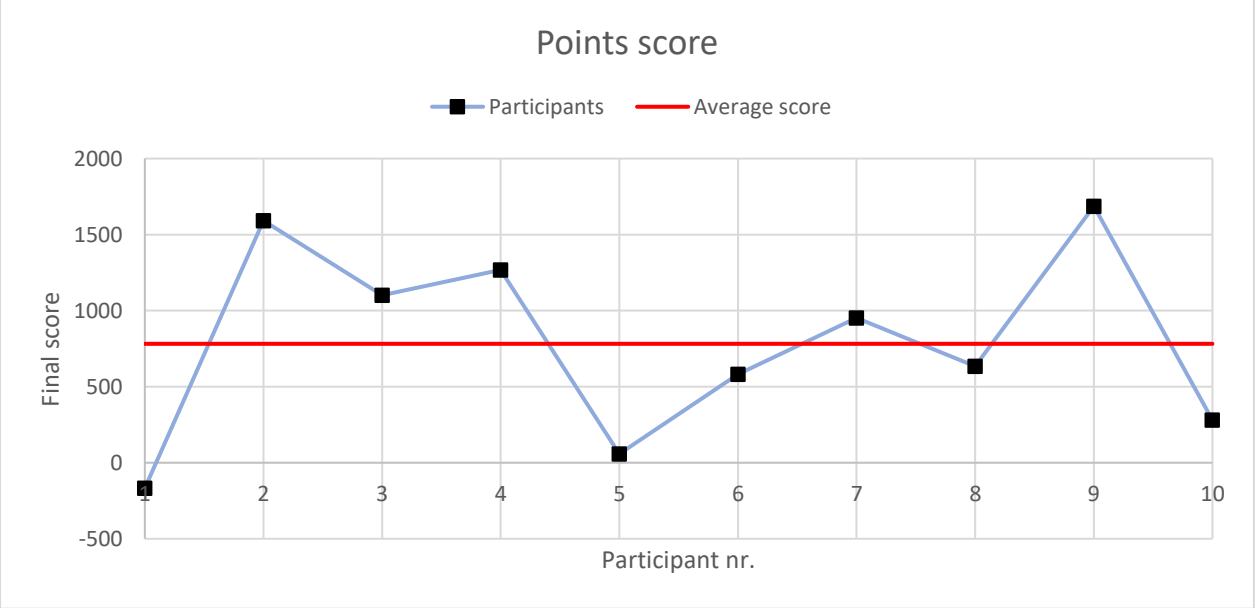


Figure 7

Average score and mean scores of the speed sub-score

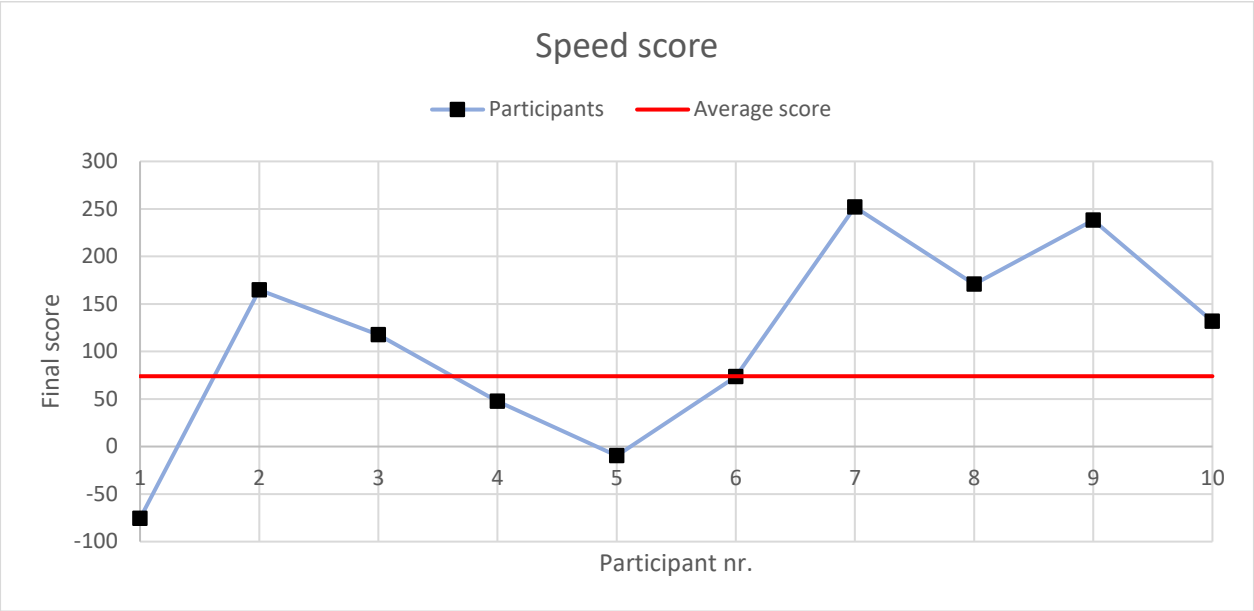


Figure 8

Average score and mean scores of the velocity sub-score

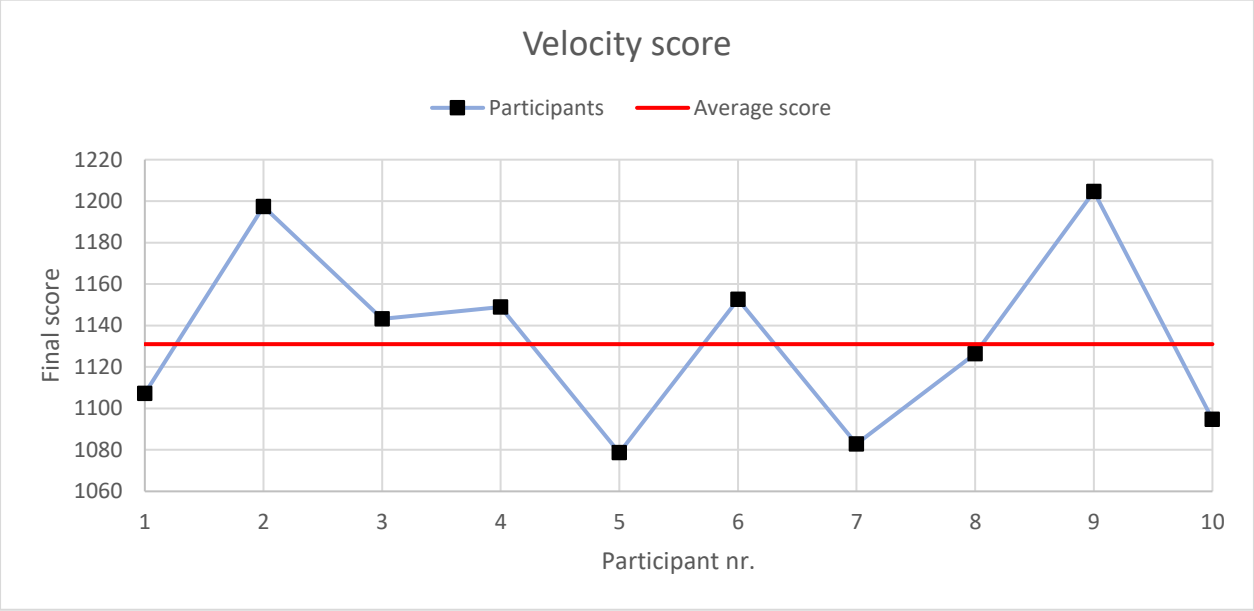


Figure 9

Average score and means scores of the total score