MASTER'S THESIS

The effects of scaffolding and feedback adaptive to the characteristics of the learner on learning progress and performance for the purpose of personalized learning

Van Mourik, Daniëlle

Award date:
2020

Open Universiteit
www.ou.nl

The effects of scaffolding and feedback adaptive to the characteristics of the learner on learning progress and performance for the purpose of personalized learning

De effecten van scaffolding en feedback aangepast aan de karakteristieken van de lerende op de leervoortgang en leerprestatie ten behoeve van gepersonaliseerd leren

Danielle van Mourik

Master Onderwijswetenschappen
Open Universiteit

Datum: 10 juni 2020
Begeleider TNO: Dr. Karel van den Bosch
Begeleider OU: Prof. Dr. Hendrik Drachsler

Table of content

The effects of scaffolding and feedback adaptive to the characteristics of the learner on learning progress and performance for the purpose of personalized learning

Danielle van Mourik

Samenvatting

Lerenden kunnen op aan aantal manieren van elkaar verschillen. Maar in het huidige onderwijs wordt er zelden rekening gehouden met verschillen, waardoor opleidingen, cursussen en trainingen vaak voor alle lerenden hetzelfde zijn. Verschillen tussen individuen bieden juist aanknopingspunten om een leertraject te personaliseren. Er zijn aanwijzingen in de literatuur dat de afstemming van ondersteuning en feedback op de behoeften van de lerende positieve effecten heeft op de kwaliteit en het tempo van leren. Deze studie wil bijdragen aan deze onderzoekslijn door een combinatie van leerstrategieën te onderzoeken die zich aanpassen aan de kenmerken van de lerende.

Het doel van deze studie is het toetsen van de effectiviteit van een gepersonaliseerd leerprogramma als combinatie van de aanpassing van (a) de moeilijkheidsgraad van oefeningen, en (b) de aard van de feedback. Beiden worden aangepast aan het prestatieniveau van de lerende. De leertaak in deze studie is het spel “Space Fortress” (Agarwal et al., 2018; Mané & Donchin, 1989). De effecten van gepersonaliseerde leerstrategieën worden onderzocht op de leervoortgang en het leerresultaat, door de resultaten van een gestandaardiseerd leerprogramma (niet gepersonaliseerd) te vergelijken met die van een gepersonaliseerd leerprogramma. De duur van het leertraject bedroeg vijf uur verspreid over twee weken. Er is een quasi-experimenteel pretest - training - posttest controlegroeponderzoek uitgevoerd onder veertig participanten die willekeurig zijn toegewezen aan de controle conditie (gestandaardiseerde leerprogramma) en de experimentele conditie (het gepersonaliseerde leerprogramma). De gemiddelde leeftijd van de deelnemers is 24 jaar. Participanten werden geworven via de proefpersonenbank van TNO.

Voorafgaand aan het leerprogramma is de self-efficacy van de participanten gemeten met de Motivated Strategies for Learning Questionnaire (Pintrich, Smith, García, & McKeachie, 1991) en hebben ze de aiming taak pretest uitgevoerd. Na de voltooiing van het leerprogramma is motivatie met de Intrinsic Motivation Inventory (Deci & Ryan, 1982) gemeten en de gepersonaliseerde leerervaring met de Personalized Learning Environment Questionnaire (Waldrip et al., 2014) en is de aiming taak posttest uitgevoerd. De leervoortgang is gemeten door middel van de prestaties (pre, mid, post) op de deeltaken, de leerprestatie is gemeten door middel van de prestatie op het complete Space Fortress spel (Frederiksen & White, 1989) en de verklarende factor voor de prestatie op het complete Space Fortress spel door middel van de aiming taak op de pretest (Mané & Donchin, 1989).

Uit de ANCOVA is gebleken dat deelnemers uit de experimentele conditie geen hoger prestatieniveau hadden op de nameting dan deelnemers in de controle conditie. Uit de repeated measures ANOVAs is gebleken dat deelnemers uit de experimentele conditie geen snellere voortgang hebben dan deelnemers in de controle conditie. Uit de T-test is gebleken dat er geen verschil is gevonden in hoe deelnemers in de controle conditie en de experimentele conditie de training beoordeelden, passend bij hun leerbehoeften. Uit de Multiple Regressie Analyse is gebleken dat zowel self-efficacy als motivatie geen voorspellers waren voor de prestatie op het complete Space Fortress spel.

In deze studie is er een compleet leerprogramma ontwikkeld waarin de aanpassingen volautomatisch en in real time zijn gevolgd en geanalyseerd en werden adaptaties volautomatisch in het leerprogramma doorgevoerd. In tegenstelling tot wat de literatuur aangeeft is uit deze studie gebleken dat een gepersonaliseerd leerprogramma niet tot een snellere en betere verwerving van complexe vaardigheden leidt dan een gestandaardiseerd leerprogramma. Mogelijke verklaringen die van invloed zijn geweest voor de resultaten is het gebrek aan interactieve multimedia instructies, gebruikerscontrole en informatie over de staat waarin de lerende zich in begeeft.

Key words: personalized learning, learning strategies, scaffolding, feedback, adaptive learning environment, learner characteristics, learning analytics

Summary

Learners can differ in a number of ways. However, in current education differences are rarely taken into account, which means that courses are the same for all learners. Differences between individuals are a starting point to personalize a learning trajectory. There is evidence in the literature that fitting support and feedback to the needs of the learner have positive effects on the quality and pace of learning. This study aims to contribute to this line of research by exploring a combination of learning strategies that adapt to the learner's characteristics.

The aim of this study is to test the effectiveness of a personalized learning program as a combination of adjusting (a) the difficulty of exercises, and (b) the nature of the feedback. Both are adapted to the learner's level of performance. The learning task in this study is the game “Space Fortress” (Agarwal et al., 2018; Mané & Donchin, 1989). The effects of personalized learning strategies are examined on the learning progress and learning outcome, by comparing the results of a standardized learning program (non-personalized) with those of a personalized learning program. The learning program took five hours extended over two weeks. A quasi-experimental pre-test - training - post-test control group study was conducted among forty participants randomly assigned to the control condition (standardized learning program) and the experimental condition (the personalized learning program). The average age of the participants is 24 years. Participants were recruited through the TNO database.

Before the learning program started, participants' self-efficacy was measured using the Motivated Strategies for Learning Questionnaire (Pintrich, Smith, García, & McKeachie, 1991) and the aiming task pretest was administered to participants. After the completion of the learning program, motivation was measured with the Intrinsic Motivation Inventory (Deci & Ryan, 1982) and the learning experience with the Personalized Learning Environment Questionnaire (Waldrip et al., 2014), the aiming task posttest was administered to participants. The learning progress was measured by the performance (pre, mid, post) on the learning tasks, the learning performance was measured by the performance on the complete Space Fortress game (Frederiksen & White, 1989) and the explanatory factor for the performance on the complete Space Fortress game by means of the aiming task on the pretest (Mané & Donchin, 1989).

The ANCOVA showed that participants from the experimental condition did not have a higher performance level on the posttest than participants in the control condition. The repeated measures ANOVAs have shown that participants in the experimental condition had no faster progress than participants in the control condition. The T-test showed that no difference was found in how participants in the control condition and the experimental condition assessed the training, according to their learning needs. The Multiple Regression Analysis revealed that both self-efficacy and motivation were not predictors of performance on the entire Space Fortress game.

In this study, a complete learning program was developed in which the adjustments were monitored and analyzed fully automatically and in real time, and adaptations were implemented fully automatically in the learning program. Contrary to what the literature indicated, this study has shown that a personalized learning program does not lead to a faster and better acquisition of complex skills than a standardized learning program. Possible explanations that have influenced the results are the lack of interactive multimedia instructions, user control and information about the state of the learner.

1. Introduction

In every learning situation there are learners with different characteristics and skills and knowledge levels. The pace at which they learn differs, as does the way learning material fits to their learning needs. Some learners need more support than others. However, education and learning programs are commonly administered in a standardized fashion, which means that courses are the same for all learners. Neither the learner's characteristics, nor the traits of the specific position in which the learner is placed after the training are taken into account. As non-standardized forms of learning are thought to be more effective. It is necessary to transform the current standardized forms of education into more flexible, individualized and personalized programs. Such programs can take differences between learners into account by shaping the learning process to their needs and capacities. This requires learning programs and learning environments that meet the learning needs of the individual learner and that are easily adaptable to the changing characteristics of the learner. For this, knowledge is needed about methods for the personalization of learning trajectories, the effects of personalization of learning trajectories on the quality and outcomes of learning and the experiences of learners during learning. This study will examine the effects of personalized learning strategies scaffolding and personalized feedback adaptive to the characteristics of the learner on the learning performance of the learner.

1.1 Theoretical framework

This thesis presents a study into the effects of personalized learning by dynamically adapting the complexity of the exercises and the nature of the feedback, to the competency level of the learner that examines the effects of learning strategies adaptive to the characteristics of the learner on the learning performance of the learner. The theoretical framework introduces this type of personalized learning and adaptive learning environments. The section 'Characteristics of the learner' introduces the four characteristics of a learner which can be used to personalize learning. Furthermore, the learning strategies *scaffolding* and *feedback* are introduced. This paragraph will conclude with the central research question and hypotheses.

1.1.1 Personalized learning

Personalized learning can be defined as a persisting change in performance or performance potential (learning results) that results from experience and interaction with the world (learning environment) (Driscoll, 2014), which meets the needs and preferences of an individual learner (Park & Lee, 2003; Sottilare et al., 2017; Salden et al., 2006). Personalized learning has been extensively studied in recent years and has shown to be an approach that can make learning more effective and engaging. There is research demonstrating that individualized instruction is superior to standardized one-size-fits-all teaching approaches (Vandewatere et al., 2013; Aleven et al., 2016; Park & Lee, 2003). One example of this type of research is a study by Bloom (1984). He found that one-to-one personalized human tutoring, compared to traditional classroom instruction, made learning performance increase with two standard deviations (Bloom, 1984). Personalized learning can be achieved by dynamic and real-time adaptation of the learning environment to a learner's unique combination of goals, interests and competencies and the ongoing process of shifting instruction as these conditions change. This is in contrast to standardized learning, which takes place in a conventional learning environment that does not meet the needs and preferences of an individual learner. Standardized learning often employs traditional instructional methods such as: giving explanations, giving instructions, and giving the opportunity for discussion (Smith et al., 2000).

1.1.2 Adaptive learning environment

The realization of personalized learning requires an adaptive learning environment (Aleven et al., 2016; Brusilovsky et al., 2007; Greller et al., 2012). Adaptive learning environments interactively respond to learner actions by adapting to the student's performance, needs and preferences, the so-called characteristics of the learner. These adaptations can be made over a short time span adaption in run time or over a longer time span adaptation by design (Aleven et al., 2016). The use of this information from and about learners to optimize the learning process and the learning environment is called learning analytics (Greller et al., 2012; Johnson et al., 2013). Examples of the type of data that can be used to personalize the learning environment are the scores of a student on summative or formative tests, but also the time at which the student studies and which question the student had difficulties with. Wetering (2016) distinguish two levels of data use, namely embedded and extracted. An example of embedded use of data is an adaptive learning environment that gives feedback or exercises fit to the level of the learner based on the input of the learner (combined with already acquired knowledge of the learner). Adaptive learning environments in which data is not used by the learning system but interpreted by a teacher is an example of extracted use of data (Wetering, 2016). Another form of extracted learning analytics is the use of data to improve the digital learning environment itself (Drachsler et al.., 2007; Romero & Ventura, 2007). Both embedded and extracted data use collect a lot of (types of) data about the learner that offer the opportunity to personalize learning.

1.1.3 Characteristics of the learner

As mentioned above, adaptivity requires information about learner characteristics in order to implement personalized learning strategies. This may, for example, involve information about a learner's personal, academic, social or cognitive self. Learner characteristics are important characteristics for designing and creating tailored instructions for the learner (Drachsler & Kirschner, 2012). Learner characteristics can be categorized into four categories (Drachsler et al., 2004; Brusilovsky et al., 2007; Aleven et al., 2016; Narcis et al., 2012, Sottilare et al., 2013; Vandewaetere et al., 2010): professional, conditional, informative and contextual characteristics. Professional characteristics (1) refer to knowledge, competences and attitudes directly related to the task. Professional characteristics are subject to change, as the goal of training is to improve or change these characteristics over time. An example of professional characteristics is the learning performance or achieved scores of the learner. Professional characteristics are not only important variables to determine appropriate contents and interventions upon during training, but they are also important outcome measures. That is why professional characteristics will be included in this study. Conditional characteristics (2) are characteristics of the learner that are not necessarily part of the learning objectives, yet are known to have a major impact on the learning process. They generally refer to self-efficacy, meta-cognitive abilities and motivation. Conditional characteristics can be changed or affected by training, thereby influencing effectiveness, efficiency and engagement of learning, but they are not necessarily the objective of training. Conditional characteristics can influence performance (i.e. professional characteristics), which is why conditional characteristics are included in the present study as well. Demographic characteristics (3) refer to relatively stable properties of a learner, like personality, gender, age and cultural heritage. These characteristics are not under the influence of a learning program and therefore not included in the present study. Contextual characteristics (4) refer to properties of the learning context that may be of importance when aiming for a personalized learning program. A few examples of contextual characteristics are: distractions, time pressure or external events such as stress. These characteristics are not included in the present study.

1.1.4 Learning strategies for personalized learning

As mentioned above, it is expected that by taking account of the characteristics of the learners, more efficient, effective and/or motivating learning strategies can be designed and developed for personalized learning (Drachsler & Kirschner, 2012). Learning strategies is an individual's way of organizing and using a particular set of skills in order to learn content or accomplish other tasks more effectively and efficiently (Schumaker & Deshler, 1992). Examples Since Bloom’s influential paper, many ways have been proposed to personalize the learning environments by differentiating instruction and adapting training methods to the needs of the individual learner, for example scaffolding, feedback, goal setting and personalization by human tutors (Van den Bosch et al., 2017). In real life, educational programs administer these strategies and interventions in combination, to achieve personalized learning. In scientific studies, however, the effects of the proposed interventions tend to be investigated in isolation (Sharma et al., 2014; Van de Pol et al., 2010; Resing, 2013; Peeters et al., 2011; Durlach & Spain, 2014, Serge et al., 2013; Shute, 2007). Many of these studies show positive effects of these individual interventions on learning. Studies into the effects of the integrated implementation of multiple interventions to achieve personalized learning are scarce (Van den Bosch et al., 2017). The largest study has been conducted by the RAND Corporation partnered with the Bill and Melinda Gates Foundation. It was found that of the 32 participating schools, students in personalized learning schools achieved higher grades in mathematics and reading than students in non-personalized learning schools (Pane et al., 2015). As this is one of the few studies that examined the effects of interventions that combine different personalization methods, the effects and impacts of personalized learning strategies on learning outcomes still require more empirical validation (Bulger, 2016; Pane et al., 2015). This need for more empirical research comes from the fact that personalization may be based upon various learner characteristics and that there are many ways to combine different learning strategies. In this study, a combination of personalization strategies will be examined. As there is evidence for the positive effects of personalized learning through the learning strategies of *scaffolding* (Van Merriënboer et al., 2004; Van der Pol et al., 2010) and *feedback* (Serge et al., 2013; Shute, 2007; Tabuenca et al., 2015), these learning strategies are included in the present study.

1.1.4.1 Scaffolding

Scaffolding is support tailored to the needs of the learner (Sharma et al., 2014; Van de Pol et al., 2010; Driscoll, 2014). Scaffolds are added during instruction within the learning task to provide the optimal level of guidance and the right kind and amount of support. The scaffolds are modified throughout a task according to a learners' progress on the task. When the learner attains the skill at a sufficient level of mastery, the scaffolds should be faded in order to remain effective (Van Merriënboer et al., 2018; Van Merriënboer et al., 2004; Zainuddin et al., 2016) (see Figure 1).

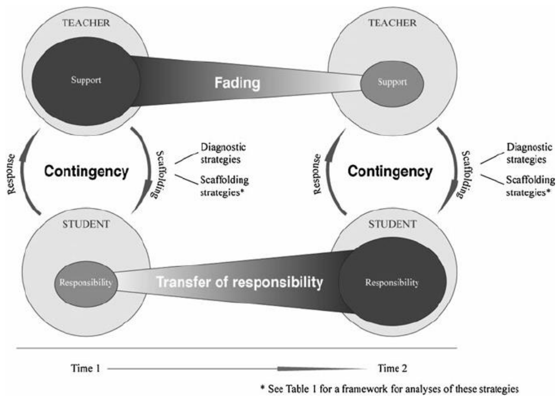

Figure 1. Scaffolding

An example of scaffolding is when the teacher gives students a simplified version of a lesson and then gradually increases the complexity or difficulty over time. Each new learning task should contain tasks and exercises that are in the zone of proximal development of the learner, meaning the difference between what a learner can do without help and what a learner can do with help (Van Merriënboer et al., 2004). When learners start to work on a new, more difficult learning task, it is essential to give them guidance and support that diminishes as the learner acquires more expertise. Scaffolding can be personalized by building a tailor-made scaffold for a learner to start learning a task or skill and adjusting this scaffold to keep it tuned and customized to the learner's needs while learning takes place. Research has shown that scaffolding is an effective strategy to support personalized learning by tailoring scaffolds to the individual needs, emotions, cognitive states and metacognitive skills of learners (Van de Pol et al., 2010; Van de Pol et al., 2011; Sharma et al., 2014; Zainuddin, 2016). When designing scaffolds, three dimensions require consideration: *what* to scaffold (content of scaffolding), *when* to scaffold (timing of scaffolding) and *how* to scaffold (method of scaffolding) (Azevedo & Jacobson, 2008). Determining what, when and how to scaffold is dependent on the characteristics of the learner (Azevedo & Jacobson, 2008).

1.1.4.2 Feedback

Feedback is the second strategy that will be used for personalized learning in the present study. Feedback is used to inform learners about their current or overall performance, including information on what they are doing (in)correctly and/or providing suggestions and guidance that allows learners to make revisions to their own performance (Serge et al., 2013). A distinction can be made between formative and summative feedback. Summative feedback is knowledge of performance after a task, set of tasks or test (e.g., percentage of correctly solved tasks, number of errors, pass or fail) (Narciss et al., 2014). Summative feedback only deals after the completion of the learning process or after the completion of the performed task and will not be able to help the learner if the learner needs help or assistance during the process. Therefore, summative feedback is not included in this study. Formative feedback can be defined as information communicated to the learner that is intended to modify the learner's thinking or behavior for the purpose of improving learning during the learner's learning process. An example is feedback provided to the learners in web-based learning dashboards, like scores or grades in a (educational) game. These types of dashboards can support awareness and reflection of individual performance and can provide suggestions for additional learning activities or content adapted to the performance level of the learner. This way, they can have a positive impact on the learning behavior (Tabuenca et al., 2015). Formative feedback has shown, in numerous studies to improve learning (Shute, 2007). Therefore, formative feedback has been chosen as a learning strategy for this study. When formative feedback is made adaptive to the learner's needs, feedback becomes personalized and is directly bound to the personal context of the learner (Tabuenca et al., 2015). Formative feedback can be personalized by adaptive-based feedback (Serge et al., 2013). Adaptive-based feedback consists of detailed feedback that switches to general feedback as scores improved past a set criterion, but also consists of general feedback that changed to detailed feedback if performance failed to improve from the previous mission score. Detailed feedback is specific information provided to the learners regarding what tasks they are performing incorrectly. General feedback is information about errors. Feedback is validated as important for learning, yet there is still some debate concerning the most effective methods for providing it (Durlach & Spain, 2014; Shute, 2007).

1.2 The present study

This study builds on research investigating the effects of personalized learning (Bloom, 1984), the effects of an adaptive learning environment (Aleven, 2016; Park & Lee, 2003), and the effects of personalized learning strategies in isolation (Sharma et al., 2014; Van de Pol et al., 2010; Serge et al., 2013; Shute, 2007) as well in a combined way (Pane et al. 2015). The present study aims to contribute to this line of research by examining a combination of learning strategies adaptive to the characteristics of the learner.

The central question in this study is: “Does a personalized learning program lead to a better acquirement of complex skills than a standardized learning program?” In this study personalized learning involves a combination of the adaptation of: (a) difficulty level of exercises, and (b) the feedback. Both are adapted to the performance level of the learner.

In this study, Space Fortress is used as the task to be learned. Space Fortress is a complex game which involves the concurrent and coordinate use of perceptual and motor skills and conceptual and strategic knowledge, in the service of multiple goals (Frederiksen & White, 1989). Space Fortress is an appropriate task to conduct research into personalized learning, because playing the game requires a large number and variety of complex skills and acquiring mastery takes quite some time (Mané & Donchin, 1989). Although, the skills learned in Space Fortress cannot necessarily be transferred directly into real life skills, the way in which the skills are taught and developed is very similar to the general process of skill learning. By using Space Fortress as learning task, we can therefore gain more insight into how the learning process of complex skills proceeds. The following hypotheses will be examined in the present study:

1. Participants that receive personalized feedback and exercises that are adapted to their performance will show a higher performance level at the end of the training than participants that receive a standardized training program
2. Participants that receive personalized feedback and exercises that are adapted to their performance will show faster learning progress during individual learning tasks than participants that receive a standardized training program.
3. Participants that receive personalized feedback and exercises that are adapted to their performance will evaluate the training as better suited to their needs than participants that receive a standardized training program.
4. Participants' level of self-efficacy and motivation will be related positively to task performance.

2. Method

2.1 Design

This study used a quasi-experimental pretest – training – posttest control group design to study the effect of an independent variable *type of learning program*. This variable has two levels: *personalized learning* and s*tandardized learning*. Effects of type of learning program on learning were investigated with the dependent variables *learning progress* (the progress that participants make during the training), and *learning performance* (the performance at the end of training).

The participants in this study were randomly assigned to either the personalized learning program (n = 20) or the standardized learning program (n = 20). Personalization of learning was achieved by: (a) adjusting the difficulty level of the task to the performance of the learner, and (b) delivering feedback that fitted the demands of the learner. These adjustments will be further described in paragraph 2.3.3 “Personalized learning program”. Participants of the standardized learning program started with the experiment. With their data being available, the learning program was then administered to the personalized learning group. This way, the data of the control group following the standardized learning program could be used to classify the performance of participants in the personalized condition as average, below average or above average. These classifications were used to adjust the difficulty level of the next exercises for participants who were assigned to the personalized learning program. In addition to adjusting the difficulty level of exercises to the participants’ competency, the nature and specificity of feedback regarding the performance on the learning task was adapted to the learner’s need. Based on the qualification as average, below average or above average, exercises and feedback can be offered that were appropriate to the learner’s level of competence.

2.2 Participants

Forty two participants (21 male; 21 female) were randomly assigned to the control condition (standardized learning) or the experimental condition (personalized learning). The age ranged from 18 to 35 years (M = 24.1, SD = 4.3). Participants were recruited among interns of TNO and the TNO database. The inclusion criteria were to have some game experience, a normal or corrected to normal vision and no other physical limitations. Furthermore, as participants were conducting their exercises from home over the internet, participants were required to have access to a computer with specifications to run the game smoothly. After completion of this study, participants were paid 10€/h for their participation. Active informed consent was obtained from all participants. In total two participants (two in the experimental condition) dropped out the experiment for personal reasons, resulting in a sample of forty participants: twenty in the control group and twenty in the experimental group.

2.3 Materials

2.3.1 Task to be learned

The digital platform used in this study is called ‘het leerproject’. In this digital web-based platform, the game Space Fortress was offered as a learning task. Space Fortress was originally developed in the 1980s for studying the acquisition of complex skills (Frederiksen & White, 1980). Figure 2 displays the interface of this game.

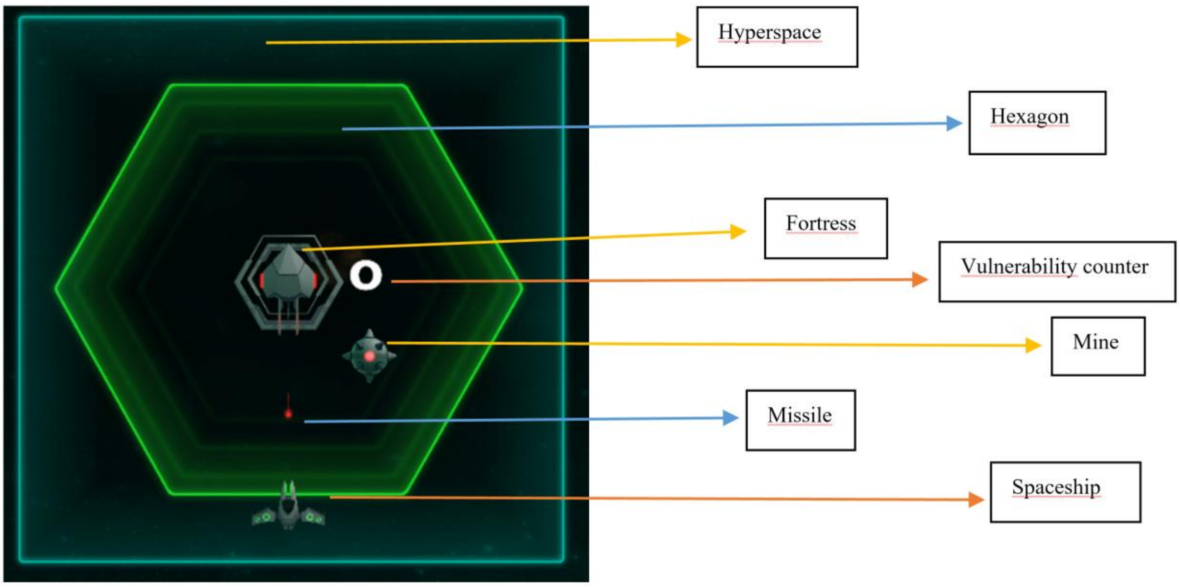

Figure 2 screenshot of the game Space Fortress

In the game Space Fortress, the user is controlling a spaceship that is navigating through space. The user has to destroy a space fortress by shooting missiles, while protecting the spaceship against damage caused by missiles that are shot from the fortress. In addition, two types of mines appear in space at set intervals, which have to be identified as either ‘friend mine’ or ‘foe mine’ by monitoring the letter that appears next to them. If the letter belongs to a pre-memorized set of three letters, it is an foe mine. If the letter is not part of the pre-memorized set, then it is a friendly mine. Friendly mines have to be energized by directly shooting them, foe mines on the other hand have to be identified as such before destroy them (Mané et al., 1989).To identify a mine as foe, the J-button must be pressed two times before the mine can be destroyed by pressing the spacebar. The interval between the two J-button presses must be between 250 and 400 milliseconds. Any interval that is shorter or longer will not be effective. After each run of a Space Fortress game, participants received feedback on their results (see Figure 3) by presenting the scores: points (number of points achieved, which can be earned by destroying the fort and / or mines), control (maneuvering the ship inside the hexagon and the playing field without getting into “hyperspace” (= outside the playing field), velocity (speed on which is flown, with a lower speed there will be a better score), speed and total score.

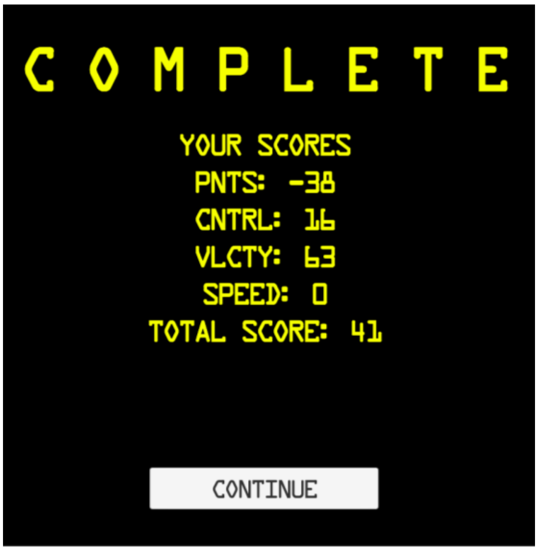

Figure 3. Feedback of results shown after each exercise and testrun of the Space Fortress game.

2.3.2 Learning program

For this study a standardized learning program was developed that supports participants in learning the Space Fortress task. This learning program is based on previous research (Oudbier, 2019), literature (Frederiksen & White, 1989; Mané et al., 1989) and a pilot study to experimentally test for user playability and game difficulty (Van Dijk, 2020). Three components were taken into account when developing the standardized learning program, namely (1) the skills to be taught, (2) the game settings and (3) an applied instruction model based on the Direct Instruction (DI) model (Van Dijk, 2020). This standardized learning program does not use personalization. This means that all participants administered all exercise and test runs in the same complexity level and receiving general feedback, which is not adapted to the learner’s needs or learning progress.

2.3.2.1 Skills to be learned

In order to perform the game two main competencies (Frederiksen & White, 1989) have to acquired, namely, (1) hitting the fortress without being hit by the fortress and (2) detecting and destroying mines (Frederiksen & White, 1989). In order to perform the two main competencies different basic skills are essential. To acquire these basic skills, 11 learning tasks were formulated, see Table 1. More detailed information on the content of the learning tasks is presented in Appendix 1.

Table 1
Description of the learning program consisting of 11 learning tasks and their connection to the two main competencies

Learning task	Total number of runs (exercise and testruns*)	Total time per learning task (min:sec)	Performance indicator	Game settings for the purpose of scaffolding	Two main competencies: Hitting the fort without being hit by the fort	Two main competencies: Detecting and destroying mines
1.Controlling the ship	22	16:30	Control	Speed of the ship	X	X
2.Firing missiles	16	12:00	Proportion shots hit = (Shots fired – shots missed) / Shots fired	the speed of the missiles fired by the fort	X	X
3.Destroying the fortress	14	10:30	Fortress destroyed	Speed of the ship	X	X
4.Stopping the ship	14	10:30	Ship stopped		X	X
5.Flying at a low velocity	14	10:30	Velocity	Speed of the ship	X	X
6.Identifying friend/foe mine	1	00:45	Multiple choice: True/false			X
7. Tagging stationary foe mines (fortress does not fire back)	16	12:00	Proportion tagged mines = Mine interval correct / (Number of defeated foe mines + number of Non-defeated foe mines)	The number of foe mine letters		X
8. Destroying stationary mines (fortress shoots back)	16	12:00	(Defeated foe mines + energized friendly mine) / (Non-defeated foe mines + defeated foe Mines + Energized + non-energized mines + ship damaged by mine)	The number of foe mine letters		X
9. Destroying moving mines	15	11:15	(Defeated foe mines + energized friendly mine) / (Non-defeated foe mines + defeated foe Mines + Energized + non-energized mines + ship damaged by mine)	The speed of the mines		X
10. Speed of destroying mines	18	13:30	Speed	The speed of the mines		X
11 Full Space Fortress game	24	18:00	Score	The speed of the ship		X

**every learning task consisted of 3 test runs (pre, mid, post)*

x connection between learning task and main competency

2.3.2.2 Game settings

The game settings of the Space Fortress task influence the complexity of the game. In developing the learning program, prior to the present study, different game settings were tested in order to develop a game that was playable for participants with a wide range of game competency. In the pilot study

(Van Dijk, 2020) three elements were found to have a major influence on the perceived difficulty of Space Fortress:

1. The maximum speed of the spaceship
2. The speed at which the missiles are fired by the fort
3. The speed at which the mines move

See Table 1 for the used game settings.

2.3.2.3 Standardized learning program

First, a standardized learning program was developed based on the direct instruction model for all learning tasks that were included in the learning program (Van Dijk, 2020). The standardized learning program gives a clear instruction based on concrete learning objectives, is teacher-driven and instructions can be checked whether it has been understood, for example by asking questions. During the instruction, learning activities can be used to better understand the material (Van Dijk, 2020). Below, the elements that were included in every individual learning task of the standardized learning program are described. The same elements were used in the personalized learning program, only then in such a wat that the complexity of the task and the feedback were adapted to the learner's competency level (see paragraph 2.3.2.4 for a description of the adaptations in the personalized learning program). The following elements were used in every learning task:

- Evaluation of what was learned in the previous task. Every learning task started with a short evaluation on what was learned in the previous learning task.
- Description of the learning goals: a description of the skills to be acquired was presented in the instruction of each learning task.
- The instruction. Task instructions were written based on the learning goals.
- Practicing. After each instruction, knowledge and skills were practiced in Space Fortress games that were suitable to practice the skills described in the learning goals. See Table 1 for the total amount of exercise and test runs per learning task.
- Check the understanding of instructions. After the instructions of the learning task, questions were asked about the given instruction to check whether the learner understood the instruction.
- The evaluation at the end of a learning task. Each learning task was concluded with a preview of the skills that would be learned in the next learning task.
- Feedback. Feedback was provided after the second and third testrun. The feedback was general in nature and included only general encouragement and general hints. By default, performance scores were presented by the game after each exercise- and testrun (see Figure 3).

2.3.3 Personalized learning program

The learning program was personalized by (1) scaffolding by adapting the task complexity to the competency of the learners, using one of three difficulty levels and (2) adaptive-based feedback.

Competent learners were assigned exercises of above-average difficulty; averagely competent learners were assigned exercises of average difficulty; and low-competent learners were assigned exercises of below-average difficulty. For each learning task, the competency of an individual learner was determined by comparing its performance to the average performance of the standardized learning group, who received all exercises on average difficulty level. If a participant's performance on a learning task was within half a standard deviation from the standard group's average, then the participant was considered to have average competency. A participant with a performance of more than +0.5 sd above the average of the standardized group was considered a competent learner; a participant with more than -.5 sd below the average of the standardized group was considered a low-competent learner.

For each learning task, a participant's competency was administered at the beginning and halfway the series of exercises, using a test run at standard difficulty level. This task run was preceded with an announcement that a test run at standard difficulty level would be administered. Performance on the test run was used to determine the difficulty level of the practice runs that followed. This way, the difficulty level at which participants practiced was not fixed, but could change twice during a learning task. Table 2 presents the possible transitions in task difficulty for participants in the personalized learning program.

Participants' task performance on the 11 learning tasks was evaluated in the final test run of that specific learning task. This final test run was again, as all test runs, administered at standard difficulty level to all participants irrespective of their competency level in order to facilitate comparison of performance between participants of both learning programs. Feedback was provided to participants during every possible transition between difficulty levels (i.e. twice during each learning task), feedback was offered tailored to the specific transition in difficulty level of the participant. The feedback was personalized by adjusting it to the assigned competency level of the participant, by addressing possible changes in the competency level (e.g., a transition from low competency to average, or from low to high competency) and general or more detailed feedback about the strategy, depending on the difficulty level the participant was assigned to. That is: participants whom were assigned practice runs at an easy level, would received detailed feedback on how they should use a strategy in order to reach good performance; participants whom were assigned practice runs at an average level would received more general feedback on what strategy they should use to reach good performance (but now how that strategy should be used); and participants whom were assigned a difficult level, would receive no feedback on strategy use.

Table 2
Possible transitions for participants of a personalized learning trajectory.

	Difficulty level of exercises prior to next (re-)assessment		
Difficulty level of exercises following (re-)assessment	Easy	Normal	Difficult
Easy	e-e	n-e*	d-e
Normal	e-n	n-n*	d-n
Difficult	e-d	n-d*	d-d

* *After the first test run, only these transitions are possible, because all participants start a learning task at standard level.*

2.3.4 Measurements

This paragraph describes the skills and knowledge measured for each outcome variable and the research question for which the variable was used. The following outcome variables were used: aiming task performance, the performance on the full SF game, performance on the learning tasks (learning progress), motivation, experiences of personalized learning and self-efficacy.

2.3.4.1 Aiming task

The aiming task can be considered a simplified version of the complete version of Space Fortress. In the aiming task, the spaceship is stationary positioned in the center of the screen. The aiming task measured the ability to swiftly rotate the ship towards the fortress and to destroy it by firing at it. Earlier research has found that performance on the aiming task is a good predictor of the performance on the full Space Fortress game (Mané & Donchin, 1989). In the present study, performance on the aiming task was therefore related to the performance on the full Space Fortress game. In addition, pretest aiming task performance was included in the analysis examining the difference between groups in performance on the full Space Fortress game at the posttest, to control for any differences between groups on learning progress.

2.3.4.2 Performance on the full Space Fortress game

Performance on the full Space Fortress game measures the two main Space Fortress competencies: the ability to hit the fortress without being hit by the fortress and to detect and destroy mines as fast as possible (see §2.3.2.1). The full Space Fortress game outputs the variable ‘Score’, which is the game’s indicator for overall performance. The average score on the last four test runs was used as indicator of performance on the full Space Fortress game. Performance on the full Space Fortress game was used for examining differences in learning performance between the standardized and personalized learning program.

2.3.4.3 Learning progress on the individual learning tasks

For each of the 11 learning tasks, a (constructed) performance indicator was defined that measured the skill(s) addressed in that specific learning task, see Table 1. This performance indicator was used for analysis of learning progress and learning performance on the individual learning tasks, after following the personalized compared to the standardized learning program.

2.3.4.4 Self-efficacy

The subscale self-efficacy of the MSLQ was used as a measure of self-efficacy. This subscale consists of 8 items and has been validated in previous studies (e.g. Pintrich, Smith, García, & McKeachie, 1991). The questionnaire was translated into Dutch and adapted to the context of the present study. The questionnaire was administered to the participants before their start of the learning program. The dataset of the present study was used to calculate reliability estimates (Cronbach's alphas) in order to verify whether the questionnaires are a reliable measure of the participants' self-efficacy. The total mean of the 8 items of the questionnaire had an excellent level of internal consistency, $\alpha = .97$, see Table 3.

2.3.4.5 Motivation

The Intrinsic Motivation Inventory (IMI) was used as a measure of motivation. The subscales interest, competence, pressure and effort were used, see Tale 3. The IMI has been validated in earlier studies (e.g. Ryan, 1982). The 23 items of the IMI were translated into Dutch and adapted to the context of the present study. The IMI was administered to the participants after finishing the learning program. The dataset of the present study was used to calculate reliability (Cronbach's alphas) in order to verify whether the questionnaires are a reliable measure of the participants' motivation. The total mean of the 23 items of the questionnaire had an excellent level of internal consistency ($\alpha = .77$), see Table 3

2.3.4.6 Experienced personalization of learning

The Personalized Learning Environment Questionnaire (PLQ) was used to measure the experiences of personalized learning. The subscales emotional, cognitive, individual assessment, congruence, transparency and academic efficacy were used, see Table 3. The PLQ has been validated in earlier studies (e.g. Waldrip et al., 2014) The 17 items of the PLQ were translated into Dutch and adapted to the context of the present study. The PLQ was administered to the participants after finishing the learning program. The dataset of the present study was used to calculate reliability estimates (Cronbach's alphas) in order to verify whether the questionnaires are a reliable measure of the participants' self-efficacy. The total mean of the 17 items of the questionnaire had an excellent level of internal consistency, α = .89, see Table 3

Table 3

MSLQ, IMI, PLQ (sub)scales, number of items and reliability

Scale	*N of Items*	*Cronbach's Alpha*
MSLQ	22	.92
Self-efficacy	8	.97
IMI	23	.77
Interest	7	.88
Competence	6	.93
Pressure	5	.81
Effort	5	.70
PLQ	17	.89
Emotional	4	.93
Cognitive	4	.66
Individual Assessment Congruence	3	.78
Transparency	3	.88
Academic Efficacy	3	.93

2.4 Procedure

All participants started with an introductory session at TNO, location Soesterberg. Participants were informed about the general purpose of the study and received practical details about their participation. They were asked to sign an informed consent form. After the introductory talk, the aiming task was administered to participants. Subsequently, participants filled out the self-efficacy scale of the MSLQ questionnaire. In total, the introductory session took approximately 30 to 45 minutes. The subsequent parts of the study were presented via a web-based research portal over the internet. This enabled participants to follow the learning programs independently, on their own computer, at their own pace, from a self-chosen location. It took them approximately four to five hours to complete the learning program. After completing the learning program participants filled out the IMI and PLQ questionnaires. This concluded their participation.

2.5 Analyses

All statistical analyses were performed with IBM SPSS version 25. Learning task 6, identifying friend/foe mines is reported, but was excluded from the analysis as the performance score were obtained using a multiple choice test for both given under the same conditions (no personalization), and therefore no performance differences were expected. To examine the effects of the personalized learning program on posttest performance on the SF task an ANCOVA was conducted with the performance on the aiming task as covariate. By including this variable as a covariate, possible effects of personalization were corrected for initial differences in task performance. To examine the effects on the learning progress over time, repeated measures ANOVAs were conducted with type of learning program as between-subjects factor and with scores on the runs in the beginning, during and at the end of each learning task as dependent variable. To examine whether the type of learning program affects how participants evaluate the training as fitting their learning needs, an independent T-test was conducted with type of learning program as between-subjects variable and the score on the PLQ as dependent variable. To examine whether participants' self-efficacy and motivation predict their task performance, the subscale Self-Efficacy and the IMI-scale were entered into a Multiple Regression Analysis with SF post performance as dependent variable.

3. Results

3.1 Exploratory data analysis

Mean scores on the aiming tasks performance on the pre- and posttest, performance on the full Space Fortress game, self-efficacy, motivation and experiences of personalized learning for the two conditions are presented in Table 4.

Table 4

Descriptive Statistics Aiming task pre- and posttest, performance on the full SF game, Self-efficacy, motivation and experience of personalized learning

	Type of learning program							
	Personalized (*n* = 20)				Standardized (*n* = 20)			
	M	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*
Aiming pre	14.30	7.98	2.00	30.00	14.75	7.92	0.00	26.00
Aiming post	20.55	8.53	0.00	33.00	21.05	6.99	11.00	34.00
Performance on the full SF game	541.66	398.81	-254	1111.2	592.18	218.13	161.25	909.75
Self-efficacy	5.71	0.53	4.50	6.63	5.17	1.49	1.88	7.00
Motivation	3.99	0.41	3.09	4.73	3.77	0.51	2.36	4.64
Experiences of personalized learning	4.98	0 .94	2.94	6.35	4.98	0.97	3.47	7.00

In general, scarce to weak correlations were found between the variables self-efficacy, motivation and experiences of personalized learning, aiming task and performance on the full Space Fortress game, see Table 5. Performance on the full Space Fortress game was not significantly related to self-efficacy r = .278, p = .082, motivation, r = .207, p = .200 and experiences of personalized learning, r = .133, p = .415. Performance on the full Space Fortress was significantly related to the aiming task r = .533, p = <.001. Self-efficacy was not significant related to the aiming task r = .206 , p = .202. Self-efficacy was significantly related to motivation, r = .589 , p = <.000, self-efficacy was significantly related to the experiences of personalized learning, r = .396 , p = .011 and motivation is significantly related to the experiences of personalized learning, r = .457 , p = .003.

Table 5
Correlation table of the variables self-efficacy, motivatie, experience of personalized learning, performance on the full Space Fortress game and aiming task

	IMI	PLQ	Performance on the full Space Fortress game	Aiming task
Self-efficacy	.589***	.386*	.278	.207
IMI		.457**	.207	.088
PLQ			.133	-.147
Performance on the full Space Fortress game				.533***

* p < .05, ** p < .01, *** p < .001

3.2 Effect of personalized learning on learning performance

An analysis of covariance (ANCOVA) was performed with post-test level of performance on the full Space Fortress game as dependent variable and aiming task performance at pretest as covariate, to examine whether participants in the personalized condition had a higher performance after completing the learning program than participants in the control condition. The mean scores on the aiming task pretest are quite the same as well as the personalized condition (M = 14.30, SD = 7.98) as the control condition (M = 14.75, SD = 7.92). The aiming task significantly predicted posttest performance level (F(1, 37) = 14.68, p = < .05). The effect of type of learning program on performance level, after controlling for the effect of the aiming task, was not significant (F(1, 37) = .219, p = .642). This result suggests that participants who received personalized feedback and exercises that were adapted to their performance level did not show a higher performance level at the end of the training than participants that received a standardized training program.

Table 6

Descriptive Statistics Task progression (pre, mid, post) of the learning tasks

		Type of learning program							
		Personalized (n = 20)				Standardized (n = 20)			
		M	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*
Leertaak 1*									
Control the ship	Pre	255.15	24.95	-715.00	270	196.20	224.65	194.00	270.00
	Mid	258.10	22.62	-81.00	270	208.45	106.01	193.00	270.00
	Post	167.80	96.15	-160.00	270	185.70	114.03	-29.00	270.00
Leertaak 2*									
Firing missiles	Pre	.573	.274	-67	1.00	.441	.363	.00	1.00
	Mid	.394	.488	-.93	1.00	.490	.405	-1.00	.95
	Post	.443	.412	-.18	.70	.085	.240	-.43	.92
Leertaak 3**									
Destroy Fort	Pre	2.00	1.62	.00	9.00	2.90	2.05	.00	5.00
	Mid	1.65	1.57	.00	8.00	2.65	2.11	.00	5.00
	Post	1.00	1.17	.00	10.00	1.80	1.95	.00	4.00
Leertaak 4**									
Stopping ship	Pre	3.20	1.40	.00	6.00	3.65	1.85	.00	5.00
	Mid	3.05	1.31	.00	7.00	3.75	2.00	.00	5.00
	Post	1.75	1.70	.00	6.00	2.30	1.95	.00	5.00
Leertaak 5**									
Moving the ship									

at low velocity									
	Pre	280.00	50.68	7.00	315.00	267.40	77.00	147.00	315.00
	Mid	253.40	66.66	21.00	315.00	263.20	73.81	119.00	315.00
	Post	295.40	36.45	147.0	315.00	286.30	57.18	217.00	315.00
Leertaak 6									
Multiple choice									
Identify									
friend/foe mines									
MC 1		.55				.80			
MC 2		.85				.90			
MC 3		.95				.90			
MC 4		.80				.75			
MC 5		.95				.85			
MC 6		1.0				.90			
MC 7		.90				1.00			
MC 8		1.0				.90			
MC 9		.95				1.00			
MC 10		1.0				1.00			
Leertaak 7**									
Tag foe mines									
(fort does not									
shoot)									
	Pre	.985	.261	.00	1.33	.939	.440	.40	1.50
	Mid	.804	.490	.00	2.00	.850	.457	.00	1.50
	Post	.310	.468	.00	1.00	.288	.380	.00	1.50

Leertaak 8** Tag foe mines(fort shoots)									
	Pre	.886	.105	.00	1.00	.743	.228	.67	1.00
	Mid	.885	.112	.40	1.00	.798	.148	.60	1.00
	Post	.708	.273	.00	1.00	.714	.317	.17	1.00
Leertaak 9 Destroy mines									
	Pre	.670	.322	.00	1.00	.768	.267	.00	1.00
	Mid	.739	.228	.00	1.00	.676	.251	.33	1.00
	Post	.730	.208	.00	1.00	.663	.340	.25	1.00
Leertaak 10 Destroying mines as fast as possible									
	Pre	85.40	109.47	-200.0	233.00	121.30	115.58	-200.0	228.00
	Mid	122.90	88.04	-1.00	220.00	112.70	78.15	-15.00	247.00
	Post	120.05	63.84	-150.0	278.00	85.55	122.63	-14.00	220.00
Leertaak 11 Full SF performance									
	Pre	618.50	235.30	-315.0	1259.0	653.60	437.04	145.00	1064.0
	Mid	548.60	291.04	-265.0	1152.0	567.25	386.89	19.00	1071.0
	Post	573.35	306.12	-115.0	1178.0	567.10	355.13	-232.00	1031.0

** significant interaction effects between personalized and control condition on progress between mid- and posttest*
*** significant main effects between mid- and posttest*

3.3 Effects of personalized learning on learning progress

To examine whether participants in the personalized condition showed faster learning progress during individual learning tasks than participants in the control condition, repeated measures analyses of variance were conducted for each individual learning task, with the scores on the runs in the beginning, during and at the end of each learning task as outcome variables. These outcomes are presented in Table 6. In the text only the significant outcomes are reported. For learning task 1, controlling the ship, Mauchly's test indicated that the assumption of sphericity had been violated, X^2 (2) = 20.25, *p* = .00, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε = .70). The interaction effect between condition and progression in learning task 1 was significant (*F*(1.40, 53.46) = 2.87, *p* = .03). A difference between the control and experimental condition was found between mid- and posttest, (*F*(1, 38) = 4.37, *p* = .04), with higher progress for the personalized condition. The interaction effect between condition and progression in learning task 2, firing missiles, (*F*(2, 76) = 4.81, *p* = .01) was also significant. A difference between the control and experimental condition was found between mid- and posttest (*F*(1, 38) = 7.99, *p* = .00), with a decline for the control condition. It is noted that decreasing scores between the pre- and posttest within the learning tasks were found, while an increasing score was expected. Main effects of progress were found in the following learning tasks:

Learning task 3, destroying the fortress, (*F*(2,76) = 16.82, *p* <.001), between mid- and posttest (*F(*1, 38) =, 13.99 *p* = .001)

Learning task 4, stopping the ship, (*F*(2,76) = 16.82, *p* <.001), between mid- and posttest (*F*(1, 38) = 21.91, *p* = .000)

Learning task 5, flying at low velocity, (*F*(2,76) = 15.13, *p* <.05), between mid- and posttest (*F*(1, 38) = 7.52, *p* = .009)

Learning task 7, tagging stationary foe mines, (*F*(2,76) = 33.67, *p* <.001) between mid- and posttest (*F*(1, 38) = 31.79, *p* = .000)

For learning task 8, tagging foe mines (the fort will shoot) Mauchly's test indicated that the assumption of sphericity had been violated, X^2 (2) = .605 *p* = .000, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε = .72), a main effect of progress was found (*F*(1.43, 54.47) = 4.82, *p* = .021), between mid- and posttest (*F*(1, 38) = 8.30, *p* = .021).

3.4 Participants' evaluation of the learning program

An independent samples T-test was performed to examine whether participants who received personalized feedback and exercises adapted to their performance level evaluated the training as better suited to their needs than participants that received a standardized training program. A non-significant effect of condition on experiences of personalization was found (*t* (38) = .274, *p* = .781). No difference was found in how participants in the control condition and the experimental condition evaluated the training as appropriate for their learning needs.

3.5 Relation of motivation and self-efficacy with learning performance

To examine whether participants level of motivation and self-efficacy was positively related to the performance on the full Space Fortress game, a Multivariate Regression Analysis was performed. The mean scores on self-efficacy are quite the same as well as the personalized condition (M = 5.71, SD = 0.53) as the control condition (M = 5.17, SD = 1.49). The mean scores on motivation are also quite the same as well as the personalized condition (M = 3.99, SD = 0.41) as the control condition (M = 3.77, SD = 0.51)The regression model with performance on the full SF game as dependent variable and motivation and self-efficacy as independent variables were not-significant, $F(2, 37) = 1.61$, $p = .213$. Only 0.03% of the performance on the full SF game could be predicted based on motivation and self-efficacy. Both self-efficacy (t = 1.22, p= .229) and motivation (t = .34 p = .735) were not significantly predictive of post-test SF performance.

4. Discussion

The main research question in this study was: "Does a personalized learning program lead to better learning and acquisition of complex skills than a standardized learning program?" In order to answer this question, the effects of personalization on learning were investigated. Personalization was achieved by (a) adjusting the difficulty of exercises to the competency of the learner and (b) providing feedback that fitted the transitions of exercises for the specific learner. The learning task in this study was the game "Space Fortress". The effects of personalization on learning progress and learning outcome were assessed, by comparing the results of a standardized learning program (non-personalized) to those of a personalized learning program. This chapter discusses the results of the five research questions, addresses the strengths and limitations of the study and provides suggestions for follow-up research.

4.1 Effects of personalization on learning and performance

Earlier research has shown that personalized learning in an adaptive learning environment with personalized learning strategies has positive effects on learning and performance (Aleven, 2016; Bloom, 1984; Park&Lee, 2003; Sharma et al., 2014; Serge et al., 2013; Shute, 2007; Van de Pol et al., 2010). We therefore expected that participants who received feedback and exercises that fit their competency level would show better learning during the program and would have a higher performance level at the end of the training than participants who received a standardized training program. However, no evidence was found for the effects of a personalized learning program on learners’ progress, nor for their achieved performance level at the end of training. These results are in contradiction with the aforementioned studies and also with that of Pane et al. (2015), one of the few scholars who examined the effects of combining various personalization strategies, finding a positive effect of this personalization on the learning performance of students. It is hard to tell why we did not find similar results in the present study. There are indications that the web-based learning environment that was used in our study failed to invoke sufficient engagement in the learners and that in the end this prevented any effects of personalization to become manifest. According to Sottilare et al. (2017) three aspects are of importance in the design of an engaging learning environment namely: (1) interactive multimedia instruction level, (2) user control and (3) the state of the learner. Following this idea, there are several possibilities for why our learning program failed to bring about the expected results. These will be discussed in the following paragraphs.

Interactive multimedia instruction

According to Sottilare et al. (2017) the interaction between the content and the learner brings about learning. This interaction can be categorized into four levels, ranging from low interaction (level 1), for example found when using reading material, to high interaction (level 4), which is for example the case in a fully immersive virtual simulation. The content used in our study can be considered to be at the lower end of the scale, as besides exercises also reading material was used to introduce the learning tasks and to explain how to perform the exercise runs. More immersive and attractive types of interactions, for example an explanation spoken by a teacher, or movie-recorded instruction materials could perhaps have set the conditions for learners to become more engaged and motivated to learn. This idea is supported by results of the IMI questionnaire, that participants were not that motivated to learn the SF game.

User control

The second aspect to take into account when designing an engaging learning environment is user control (Sottilare et al., 2017). User control is thought to have a large effect on learning performance. User control can be provided in an adaptive learning system by:

- offering the learner a means to initiate/halt adaptation of the system during every phase of learning
- allowing the learner to accept, modify or reject every or any part of proposed adaptation
- enabling the learner to specify adaptation parameters
- informing the user about the proposed changes due to adaptation before actual changes take place
- giving the learner access and sole control over his/her behavior records and their evaluation (open learner model)

In our learning program, mandatory fixed 15-minute breaks were used between learning tasks. Participants who wanted to continue with the next learning task immediately were not given the opportunity to do so. Besides this, the learning environment unfortunately did not provide a 'back' button to read the instructions again if the participants wanted to. Due to the lack of interaction between the system and the participants (Aleven et al., 2016), the participants were not given the opportunity to reject or accept the indicated adjustment in a subsequent exercise. The participants were also not given the opportunity to indicate which parameter of the game should be adjusted. Neither did they have the free choice to select the next task (Van Merriënboer et al., 2004). They had to follow the changes and the given learning tasks. It can be concluded that the lack of user control reduced the chance of effects on personalization on learning in our study. This is supported by the comments of the participants that they had the preference to decide by their selves to had a break or not between the learning tasks.

State of the learner

The last aspect that is important in the design of engaging learning environment is the state of the learner (Sottilare et al., 2017). The classification of the students' situation refers to the mental or physical situation of the learner, which can influence the learning performance of the learner. In a learning process the learner goes through a development on the basis of instructions, the difficulty of the excercises and the feedback that the learner receives, which is adapted to his learning needs. The learning performance is also formed by the physical situation and/or mental situation, in which the learner finds himself. Depending on the task to be learned, an adaptive learning environment can respond to the state of the learner. In our learning environment we measured learning performance and task progress, but we do not measure other states, traits and preferences. For example, adaptation could also include preference tailoring in which the environment is adapted to the specific learner's cultural background to provide a familiar mental model for learning. This could enhance learner engagement and result in less down time during instruction. Another adaptation to improve effectiveness could include tailoring based on learner interests. In this study, we could have adapt the amount of exercise and test runs that accidentally were not accurate due to technical failures. This assumption is supported by the comments of the participants that they became 'discouraged' because of the amount of exercise runs exceeded compared to the given amount of exercises in the feedback.

4.2 Effects of personalization on how training and learning is experienced

According to Waldrip et al. (2016) learning is experienced as personalized if the learning environment demonstrates concern for, and knowledge of, students as individuals. This can be achieved by providing strategies to address their particular academic and socio-emotional needs and well-being. Therefore, it was expected that participants who received personalized feedback and exercises that were adapted to their performance, would evaluate the training as better suited to their needs than participants who received a standardized training program. However, no effects of personalization were found on how training and learning were experienced. There are indications that the lack of measuring the state of the participant with the result of not demonstrating concern for participants or the strategies to address their particular socio-emotional needs and well-being has reduced the chance of effects on personalization on learning. This is supported by the results of the IMI questionnaire, that participants were not that motivated to learn the SF game.

4.3. Motivation and self-efficacy as predictors of learning and performance

Research has repeatedly shown that motivation and self-efficacy are major factors influencing students' academic success (Arroyo et al., 2014). The review by Aleven et al. (2016) confirms this, finding that affect and motivation had a positive influence on student learning. Therefore, it was expected that the participants' level of motivation and self-efficacy in the present study would be positively related to task performance. Based on our results, the hypothesis could not be confirmed: neither motivation, nor self-efficacy was related to posttest performance. A possibility for our finding that a participant's motivation at the onset of training does not predicts its subsequent learning and performance may be due to the fact that motivation and the affective state of the participant were only measured at one timepoint, instead of frequently throughout the learning program as was done in the study of Aleven et al. (2016). Therefore, our results might not give a complete picture of the fluctuations in motivation and self-efficacy that participants experienced during training. There are indications in support for this possibility, for example quite a few participants indicated that they felt very demotivated and discouraged during the learning program. By measuring the affective state of the participants

Another reason for the finding that self-efficacy was not related to task performance may have been that self-efficacy was measured directly after participants conducted the aiming task, which is a fairly easy task to carry out. This may have been of influence on the expectations that participants formed on the learning task Space Fortress. The aiming task might have given the impression that the Space Fortress task would be an easy task possibly resulting in high expectations about their own efficacy. This is supported by the results of the Self-Efficacy subscale, revealing that participants had a quite high level of self-efficacy.

4.4 Strengths and limitations of the present study

A web-based platform that offers many possibilities is used for this purpose. An important strength of this study is the digital learning platform that was used. There are digital learning environments where a lot of data has to be collected in order to predict the learning processes of learners. The platform we used automatically and in real time monitored and analyzed the learner's performance and used this to automatically implement adaptations in the learning program. Furthermore, the web-based platform offered participants the opportunity to learn and practice in a self-chosen time and location (Sottilare et al., 2015). Besides this, the participant has been informed about the imminent changes before these were actually implemented (Sottilare et al., 2017). Finally, the learning environment includes instructions and exercise and test runs. As a result, the information collected about the learner and the learners learning process can be integrated.

The learning environment offered learning possibilities, but also had limitations. Due to technical failures, too many exercise runs were mistakenly prepared for some participants. As a result, the number of the announced runs was no longer accurate. Therefore, the participant could possibly be of his apropos and no longer had any confidence in the learning system or in his learning progress. “Would there be another exercise run or would I have completed the learning task after this exercise run,” or “am I really bad in this game that I did not get a certain level that I have to practice a lot” were comments from participants.

Furthermore, the scores decreased from mid- to post-run. A possibility to explain why these scores decreased is that the test runs were announced, which could have caused stress, because the participants wanted to perform very well. To see if there is a significant difference between the pre and midtest in the learning task, additional mixed anovas were administered. These analyzes showed that there is no significant difference between the pre- and mid-test and that there is no significant difference between the skills acquired at the time of the pre-test and the skills acquired at the time of mid-test. The assumption that there are differences in the mean score between the pre- and midpost can be rejected.

4.5 Suggestions for follow-up research

Aforementioned research supported by results of the present study and comments of the participants have shown three aspects may be of influence for developing an adaptive learning environment namely: (1) interactive multimedia instruction level, (2) user control and (3) the state of the learner. Therefore, a follow-up research can be conducted included these three aspects. Furthermore, the averages of the scores for the learning tasks showed a slight increase between the pre- and mid-tests. But instead of continuing the learning process, they seem to drop in performance from the mid to posttest of a learning task. This phenomenon may lie in the so-called cognitive load theory (Sweller, 211). Cognitive load is a situation where the learner received too many tasks simultaneously, resulting in the learner being unable to process this large amount of information. Cognitive load can be prevented by using scaffolding. In this study, we used scaffolding as complexity support. In a follow-up research scaffolding could also have been used for the number of times a task was practiced, in which the learner could have a choice. Finally, the way the skills are taught in Space Fortress is very similar to how we learn skills in general, but the skills learned in Space Fortress cannot necessarily be translated directly into real life skills. Therefore, follow-up research with a learning (web)environment investigating the effects of combined interventions that together constitute personalized learning can also be explored using another task than Space Fortress in order to acquire real life skills. The intertwining of education and technology will certainly not diminish in the coming years and education will reach both the learner and the teacher for the digital world. It is therefore of great importance to investigate which learning strategies are most effective related to personalized learning and which technological learning environments can best play a role in this.

Acknowledgements

I would like to extend my gratitude to Karel van den Bosch, Anne de Bruijn, Romy Blankendaal and Hendrik Drachsler for their feedback, supervision and/or support during my graduation process. In addition, I would like to thank the department Training and Performance Innovations of TNO for allowing me an internship.

6. References

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. In R.E. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 522-560). Abingdon, United Kingdom: Routledge.

Anand, V. (2017). Design Considerations for Personalized Learning. In: A. Chakrabarti and D Chakrabarti (Eds.), *Proceedings of the International Conference on Research into Design*. Singapore, Singapore: Sprinter

Azevedo, R., & Jacobson, M.J. (2008). Advances in scaffolding learning with hypertext and hypermedia: A summary and critical analysis. *Educational Technology Research and Development, 56*(1), 93-100

Bosch, K. van den, Peeters, M.M.M., & Boswinkel, R. (2017). *Literature review on individual learning concepts. Part A: Personalized Learning*. Soesterberg, the Netherlands: TNO.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, *13*(6), 4-16.

Brusilovsky, P., & Millán, E. (2007). *User models for adaptive hypermedia and adaptive educational systems*. In The adaptive web (pp.3-53). Springer: Berlin, Heidelberg

Bulger, M. (2016, July 22). Personalized Learning: The Conversations We're Not Having. *Data and Society*. Retrieved from: https://www.datasociety. Net

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

Cresswell, J.W., (2014). *Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research*. Upper Saddle River, N.J: Merrill.

DAOG (2016a). *Visie opleiden & individueel trainen. Leren & Ontwikkelen bij Defensie – 2030*. Den Haag: Ministerie van Defensie

Drachsler, H., Hummel, H.G.K., Koper, R. (2007). *Recommendations for learners are different: applying memory-based recommender system techniques to lifelong learning*. Paper presented at the conference SIRTEL workshop at the EC-TEL 2007. September 17-20, 2007, Crete, Greece.

Drachsler, H., Kirschner, P. (2012). *Learner Characteristics*. In (Ed.) Seel, Norbert M. 1st Edition Encyclopedia of the Sciences of Learning. Berlin: Springer.

Driscoll, M.P. (2014). *Psychology of learning for instruction*. Essex: Pearson Education Limited

Durlach, P.J., & Spain, R.D. (2014). *Framework for Instructional Technology: Methods of Implementing Adaptive Training and Education*. (Technical Report 1335). Fort Belvoir, VA: US Army Research Institute for the Behavioral and Social Sciences.

Frederiksen, J.R., & White, Y.W. (1989). An approach to training based upon principled task decomposition. *Acta Pscyhologica*, *71*, 89-146.

Goldberg, B., Brawner, K., Sottilare, R., Tarr, R., Billings, D.R., & Malone, N. (2012). *Use of evidence-based strategies to enhance the extensibility of adaptive tutoring technologies.* In the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).

Greller, W. & Drachsler, H. (2012). Translating learning into numbers: A generic framework for Learning Analytics. *Eduational Technology & Society*, *15*(3), 42-57.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, *66*(9), 641–661.

Mané, A., & Donchin, E. (1989a). Adaptive and part-whole training in the acquisition of a complex perceptual motor skill. *Acta Psychologica*, *71*, 179-196

Mané, A., & Donchin, E. (1989b). The Space Fotress Game. *Acta Psychologica*, *71*, 17-22

Narciss, S., Sosnovsky, S., Schnaubert, L., Andres, E., Eichelmann, A., Goguadze, G., Melis, & E. (2012). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education*, *71*, 56-76

Oudbier, J. (2019). *The effect of personalization on learning a complex skill* (Master's thesis, Radboud University, Nijmegen, Nederland).

Pane, J.F., Steiner, E.D., Baird, M.D., & Hamilton, L.S. (2015). *Continued Progress: Promises Evidence on Personalized learning*. Santa Monica, CA: RAND Corporation.

Park, O. C., & Lee, J. (2003). Adaptive instructional systems. *Educational Technology Research and Development*, *25*, 651-684.

Peeters, M., Bosch, K. van den, Meyer, J.J. Ch., & Neerincx, M. (2011). *Scenario-based training: Director's cut*. Conference conducted at the 15th International Conference on Artificial Intelligence and Education (AIED), Auckland, NZ.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33.

Resing, W.C. (2013). Dynamic testing and individualized instruction: Helpful in cognitive education?. *Journal of Cognitive Education and Psychology*, *12*(1), 81.

Salden, R.J.C.M., Paas, F. & Van Merriënboer, J.J.G. (2006). Personalised adaptive task selection in air traffic control: effects on training efficiency and transfer. *Learning and Instruction*, *16(*4), 350-362.

Schumaker J.B., Deshler D.D. (1992) *Validation of Learning Strategy Interventions for Students with Learning Disabilities: Results of a Programmatic Research Effort*. In: Wong B.Y.L. (eds) Contemporary Intervention Research in Learning Disabilities. Disorders of Human Learning, Behavior, and Communication. Springer, New York, NY

Serge, S.R., Pries, 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.

Sharma, S., Ghorpade, S., Sahni, A., & Saljua, N. (2014). Survey of Intelligent Tutoring Systems: A Review on the Development of Expert/Intelligent Tutoring Systems, various Teaching Strategies and Expert Tutoring System Design Suggestions. *International Journal of Engineering Research & Technology*, *3*(11).

Shute, V.J. (2007). *Focus on Formative Feedback. Research Report*. Princeton, NJ: Educational Testing Service

Smith, S.B.S., Smith, S.J., & Boone, R. (2000). Increasing Access to Teacher Preparation: The Effectiveness of Traditional Instructional Methods in an Online Learning Environment. *Journal of Special Education Technology*, *15*(2), 37-46.

Sottilare, R.A. (2015). *Challenges in Moving Adaptive Training & Education from State-of-Art to State-of-Practice*. Orlando, FL: U.S. Army Research Laboratory.

Sottilare, R.A. & Goodwin, G.A. (2017). *Adaptive instructional methods to accelerate learning and enhance learning capacity.* Orlando, FL: U.S. Army Research Laboratory

Sweller, J. (2011). Cognitive load theory. *Psychology of learning and motivation*, *55*, 37-76.

Tabuenca, B., Kalz, M., Drachsler, H., Specht, M. (2015). Time will tell: the role of mobile learning analytics in self-regulated learning. *Journal of Computers & Education*. Doi:10.1016/j.compedu.2015.08.004

Van de Pol, J., Volman, M., Elbers, E., & Beishuizen, J. (2012). Measuring scaffolding in teacher-small-group interactions. In R. M. Gillies (Ed.), *In Pedagogy: New Developments in the Learning Sciences* (pp. 151-188). Hauppage: Nova Science

Van de Pol, J., Volman, M., & Beishuizen, J. (2011). Patterns of contingent teaching in teacher-student interaction. *Learning and Instruction*, *21*, 46-57.

Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher– student interaction: A decade of research. *Educational Psychology Review*, *22*(3), 271-296.

Vandewaetere, M., Desmet, P., & Clarebout, G. (2010). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, *27*, 118 – 130.

Van Dijk, C.G. (2020). *Adviesrapport standaard leertraject voor Space Fortress*. TNO: Soesterberg, Nederland

Van Merriënboer, J.G. & Kirschner, P.A. (2018). *Ten steps to complex learning. A systematic approach to four-component instructional design*. New York: Routledge

Waldrip, B., Yu, J.J. & Prain, V. Validation of a model of personalised learning. *Learning Environ Res 19,* 169–180 (2016).

Waldrip, B., Cox, P., Deed, C., Dorman, J., Edwards, D., Farrelly, . . . Yager, Z. (2014). Student perceptions of personalized learning: Development and validation of a questionnaire with regional secondary students. *Learning Environments Research*, *17*(3), 355–370.

White, B. Y., & Frederiksen, J. R. (2005). A theoretical approach and framework for fostering metacognitive development. *Educational Psychology*, *40* (4), 211-223.

Woning, E. (2013). Whitepaper Learning Analytics: Op weg naar meer gedifferentieerd en gepersonaliseerd onderwijs. Zoetermeer: Stichting Kennisnet.

Zainuddin, N., Sahrir, M. S., Idrus, R. M., & Jaafar, M. N. (2016). Scaffolding a conceptual support for personalized Arabic vocabulary learning using augmented reality (ar) enhanced flashcards. *Journal of Personalized Learning*, *2*(1), 102-110.

APPENDIX 1 – SUBGOALS AND CONTENT OF THE LEARNING TASKS

Subgoals (Frederisken & White, 1980)	**Learning tasks**
Controlling the ship	Learning task 1 – controlling the ship In this learning task the participant must learn to control the ship by rotating the ship clockwise or counterclockwise. In addition to this, the participant tries not to fly into the hyperspace and not to touch the inner hexagon.
Firing missiles	Learning task 2 – firing missiles In this learning task the participant must learn to fire missiles. Since shots at the fort must not occur too rapidly, they must learn to fire at an optimal rate, to keep track of the shot counter (in addition to this, the participant tries not to fly into the hyperspace and not to touch the inner hexagon).
Destroying the fortress	Learning task 3 – destroying the fortress In this learning task the participant must learn to destroy the fortress by a 'double shot'. Once the participants hits the Fortress 10 times (or more), the participant can destroy it by hitting it with a double shot by two spacebar presses. The interval between the two spacebar presses must be between 250 and 400 milliseconds.
Stopping the ship	Learning task 4 – stopping the ship Participants must learn to rotate the ship so that they can apply a thrust in the direction opposite to the ship's motion to stop the ship.
Moving the ship at a low velocity by monitoring the velocity score	Learning task 5 – moving the ship at a low velocity by monitoring the velocity score In this learning task the participant must learn to control the ship at a constant angular velocity. The score for velocity is continuously updated and displayed on the instrument panel underneath the label "VLCTY".

Identifying friend and foe mines	Learning task 6 – identifying mines In this learning task participants need to learn how to monitor for the occurrence of mine identification letter and to correctly classify the mine as friend or foe on the basis of the letter presented by using the IFF display. A demo will show how to identify a mine. The participant indicates whether this is an enemy mine or a friendly mine.
Tagging of foe mines (the fortress does not shoot back)	Learning task 7 – tagging foe mines (the fortress does not shoot) In this learning task the participants learns how to tag a foe mine by pressing twice the "J"-button with an interval between the button presses of 250-400 milliseconds.
Tagging of foe mines (the fortress shoot back)	Learning task 8 – tagging foe mines (the fortress will shoot) Idem as in learning task 7, but the fortress will shoot back.
Destroying moving mines	Learning task 9 – destroying mines In this learning task participants need to learn to destroy mines with the spacebar after identifying a mine as foe or friend. A friendly mine will be destroyed by pressing once the spacebar. A foe mine will also be destroyed by the spacebar, but the participant must press first the "J" button before it can be destroyed by the spacebar. In addition to this, the participant learnt how to destroy mines without running from them. They need to learn to locate the mine while they are making the IFF response, so that they can quickly judge whether to reaim the ship in order to hit the mine or whether they can hit the mine without reaiming. They also need to learn the regions where mines are most likely to appear, given the current position of the ship.
Destroying mines as fast as possible	Learning task 10 – destroying mines as fast as possible In this learning task participants need to learn to destroy the mines as fast as possible.