Cleared for take-off: Game-based learning to prepare airline pilots for critical situations

Kuindersma, E.C.

Chapter 6

Measuring the Shuttle to Mars experience

In this chapter, we will present a small-scale qualitative study conducted to investigate how airline pilots experience the *Shuttle to Mars* game. We will answer RQ 3, which reads as follows.

RQ 3: *To what extent do airline pilots accept a game to develop essential competencies for critical situations?*

The design of the *Shuttle to Mars* game aims at developing airline pilot competencies for critical situations. It was described in Chapter 5. For the design of *Shuttle to Mars*, we attempted to apply the guidelines for serious games for competency development, presented in Chapter 3.

In Section 6.1, we will discuss why and how the game experience of the *Shuttle to Mars* game should be evaluated. In Section 6.2, we will describe the set-up of our study, including the methodology and materials. Then, in Section 6.3, we will present our findings. Subsequently, in Section 6.4, we will evaluate and discuss our findings. Finally, in Section 6.5, we will answer RQ 3 and give recommendations for further research.

6.1 Evaluating the Shuttle to Mars game experience

Training of professionals aims to improve the professional's performance on the job, which leads to benefits for the organisation. This is the top of the pyramid in Kirkpatrick's model of training evaluation [105, 106] (see Figure 6.1).

The pyramid represents the hierarchical nature of the model. The higher levels are of greater importance [17], and can be reached when the lower levels are satisfied [76]. The three top levels are based upon the reaction (i.e., the affective responses) of the participants to the quality or relevance of the training [17]. When participants are positive about the course (level 1), there is a better chance that they will learn (level 2) and

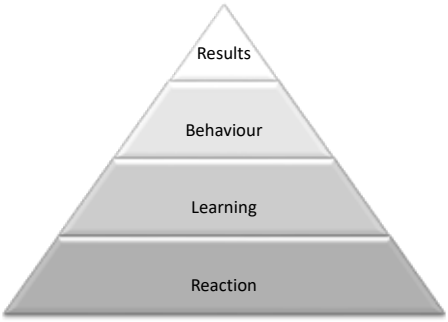

Figure 6.1: Kirkpatrick's four levels of training evaluation

change their behaviour (level 3), resulting in improvements for the organisation (level 4). However, Kirkpatrick's model is not a prescriptive model. A positive reaction alone is not sufficient.

In our research, we aim to determine whether a serious game can be used to train airline pilots to act adequately in critical situations. We investigate the reaction that airline pilots have towards a serious game that was designed specifically for that purpose.

For the organisations involved in aviation, the intended result of such a training is a further decrease in aircraft accidents that can be attributed to human error. Part of accomplishing such a reduction in accidents is to train airline pilots to change their behaviour in the cockpit. The idea is that behaviour resulting from the competencies identified in Section 5.1, will enable the pilots to act adequately in critical situations.

A change in behaviour should be the intended result of the learning that takes place in the training environment. In our research, the intended result is the development of essential competencies through playing the *Shuttle to Mars* game. To allow this learning to take place, the learners should have a positive reaction to the serious game and accept it as a valid way of developing their competencies.

A reaction of accepting the training is the first level of the training evaluation pyramid [105, 106]. It is essential for the higher levels of effect to be reached. Therefore, it is important first to assess the reaction of the players to a game. We will study their experience in order to ascertain the extent to which the learners will accept the game as a training method.

We expect that a player who gives a response of accepting the game will be more motivated to play the game and to continue the experiment. Therefore, we will take the participants' level of motivation as a measure of their positive attitude towards the game.

In the design of a serious game, attention needs to be paid to the balance between fun and learning [65, 128, 171]. Consequently, the experience of a serious game involves fun as well as learning. Both types of experience should be measured when evaluating a serious game.

In our study, we will investigate how the participants experience the *Shuttle to Mars* game. We will address their enjoyment of the game, and their expectations about developing their competencies.

6.2 Research set-up

In this section, we will describe the set-up of our study. We will start with the methodology (Subsection 6.2.1). Next, we will discuss the participants (Subsection 6.2.2). Then, we will explain our procedure for measuring the experiences with the *Shuttle to Mars* game (Subsection 6.2.3). Finally, we will discuss the materials used (Subsection 6.2.4).

6.2.1 Methodology

In our research, we aim to measure airline pilots' reaction to our *Shuttle to Mars* game, and the experience they have playing the game. For this purpose, we used (1) online questionnaires (see Appendix D) to collect qualitative data including some demographic information, as well as some quantitative data, and (2) audio-recorded semi-structured interviews (see Appendix D.5) to collect qualitative data.

For the qualitative study, we have chosen the research strategy of *grounded theory* [74]. The grounded theory approach aims to allow a theory to emerge from data about a reality that is being investigated. It uses a systematic set of procedures to inductively derive a theory from data that is collected and analysed pertaining to a specific phenomenon [186]. A grounded theory represents a reality. Therefore, it should make sense to the persons who were the object of study as well as those who practise in the same area.

6.2.2 Participants

All participants were recruited through connections with Dutch airline companies. Our research is aimed at providing experience for airline pilots with little experience. In consultation with airline pilots and a flight instructor, we quantified *little experience* as a maximum of 5 years and 2500 flight hours with an airline. After an open call for participation, initially, sixteen pilots showed interest to participate. However, at the start of the experiment, only five pilots were available to participate due to conflicting schedules. Each participant was informed about the purpose of the study and signed a consent form.

All pilots were male, with a mean age of 31.2 years ($SD = 6.3$). They were first and second officers with three different airlines. On average, they had 2.6 years of experience working as a pilot for an airline company ($SD = 1.1$). In flight hours, their experience ranged from 1000 to 2300, with a mean of 1880 hours ($SD = 870$).

One participant reported to play games daily, on average, an hour per day. Three participants said that they play games a few times per week. The last participant rarely plays games. On average, the pilots say to play games for 30 minutes per week.

In addition to the young airline pilots, we invited three experienced flight instructors to participate in the study. The procedure for the flight instructors was identical to that for the airline pilots.

All flight instructors were male. They were older than the other group ($M = 39.3$, $SD = 7.6$) and their flight hours ranged from 5000 to 9000. All instructors were captains affiliated with an airline.

Participants and instructors all received reimbursement for time and travel expenses, in the amount of €100. As an incentive to complete all parts of the study, a modern smart-watch was raffled upon completion of the experiment by way of a random draw. Pilots and instructors all had an equal chance of winning the watch.

6.2.3 Procedure

The procedure consisted of five parts, pictured in Figure 6.2. Participants had a Start session in the NLR offices in which they acquainted themselves with the game by playing the tutorial (see Subsection 5.2.2). Then, in their own time and on their own personal computers, they played the game. It was divided into three blocks. Each block consisted of approximately 1 to 1.5 hours of gameplay and, on top of that, a questionnaire. Participants had one week to complete each block. They did not have to complete a block all at once. After completion of the game, the participants were invited back to the NLR offices for the final interviews. The interviews lasted 30 to 40 minutes per participant.

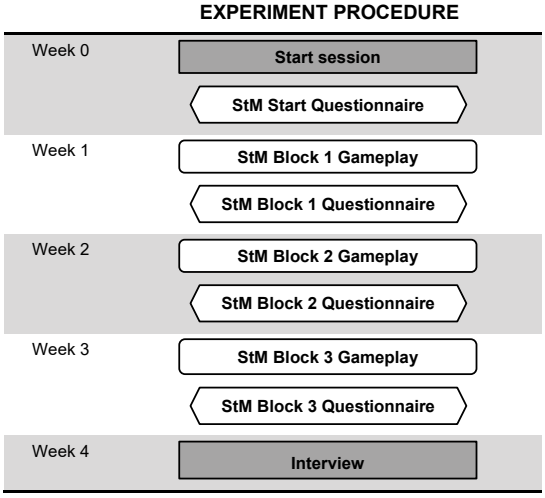

Figure 6.2: Procedure of the small-scale qualitative study for Shuttle to Mars (StM)

6.2.4 Materials

Playing the *Shuttle to Mars* game is the central part of the current study. To record the players' experiences with the game, four questionnaires and a semi-structured interview scheme were developed. Data generated by playing the game was stored in the game database for further research of in-game behaviour and learning effect. However, as our study focused on determining the acceptance of the game by airline pilots and not on finding a statistical proof of its learning effect, the game data was not used for the purpose we had in mind.

The design of the *Shuttle to Mars* game has been discussed in Chapter 5. In this subsection, we will discuss (A) the questionnaires and (B) the interview structure.

A. Questionnaires

Each participant was presented the StM Start questionnaire during the Start session, and then an StM Block questionnaire after completing each of the three game blocks. In addition to questions about the gameplay, each questionnaire consisted of a set of repeating questions and a validated tool. We used three validated tools: (1) the UEQ, (2) the IMI, and (3) the CEGEQ.

1. The UEQ is a tool to evaluate the quality of the user experience of an interactive product [164, 181].
2. The IMI is a multidimensional instrument to measure a participant's subjective experience about the activity in an experiment IMI [91, 172] (see also Chapter 4).
3. The CEGEQ is an instrument to assess the presence of elements that are necessary but not sufficient for a positive game experience [28, 29].

Repeating questions relating to motivation. Each of the StM Block questionnaires had a set of repeating questions to gauge the motivation of the participant. These questions enquired after:

1. time spent playing the *Shuttle to Mars* game and other games,
2. types of other games played,
3. feeling about and attitude towards the game and the experiment.

Repeating questions relating to gameplay and the relation to the job. Each of the StM Block questionnaires had some questions to reflect on the gameplay and its relation to an airline pilot's job.

Below, we describe the four questionnaires for (i) Start session, (ii) Block 1, (iii) Block 2, and (iv) Block 3. The validated tools included in the questionnaires will also be described. See Appendix D for complete questionnaires.

A1. StM Start questionnaire

The StM Start questionnaire gathered demographic and gaming related information. Moreover, it enquired after the participant's attitude towards innovations and their motivation to participate. As this questionnaire was answered right after the first gameplay session, questions were asked about whether it was clear how the game should be played.

A2. StM Block 1 questionnaire (with the UEQ)

The StM Block 1 questionnaire comprised the UEQ [164, 181]. The UEQ contains 26 items on six scales [181], describing distinct quality aspects, such as "boring - exciting". The six scales are as follows.

1. **Attractiveness.** Do users like or dislike the product?
2. **Perspicuity.** Is it easy to get familiar with the product? Is it easy to learn how to use the product?
3. **Efficiency.** Can users solve their tasks without unnecessary effort?
4. **Dependability.** Does the user feel in control of the interaction?
5. **Stimulation.** Is it exciting and motivating to use the product?
6. **Novelty.** Is the product innovative and creative? Does the product catch the interest of users?

Scores have to be given spontaneously and have to reflect the personal opinion of the participant. Each pair is scored on a 7-point scale. In the analysis, the scores are transformed to a range from -3 (horribly bad) to 3 (extremely good). Scores between -0.8 and 0.8 are considered neutral.

Rauschenberger et al. [164] have grouped the six UEQ scales into "*pragmatic* and *hedonic* quality" (Table 6.1). Pragmatic quality describes task-related quality aspects, viz. perspicuity, efficiency, and dependability. Hedonic quality describes non-task-related aspects, viz. stimulation and novelty. According to Rauschenberger et al. [164], attractiveness is neither pragmatic nor hedonic. Therefore, they label it as a "pure *valence* dimension" [164, 181].

Table 6.1: Categorisation of UEQ scales

Pragmatic	Hedonic	Valence
Perspicuity	Stimulation	Attractiveness
Efficiency	Novelty	
Dependability		

A3. StM Block 2 questionnaire (with the IMI)

The StM Block 2 questionnaire comprised the questions of the IMI [91, 172]. The IMI consists of twenty statements concerning six subscales (see Table 4.7 on p. 76). Agreement with each statement is scored on a 7-point Likert scale, ranging from *not at all true* to *very true*. The items of the IMI instrument are divided [172] into six subscales.

1. Interest/Enjoyment
2. Perceived Competence
3. Effort
4. Value/Usefulness
5. Pressure/Tension
6. Perceived Choice

In addition to the original IMI statements, we added three statements in the same fashion about their attitude towards gaming [197] to the StM Block 2 questionnaire. For a full list of all statements, we refer to the StM Block 2 questionnaire in Appendix D.

A4. StM Block 3 questionnaire (with the CEGEQ)

The StM Block 3 questionnaire comprised the CEGEQ [28, 29]. With the Core Elements of Gaming Experience (CEGE) framework, Calvillo-Gamez [28, 29] presents a series of elements that are indispensable for a positive game experience. According to Calvillo-Gamez, these elements are "necessary but not sufficient". If all are present, the gaming experience will not be negative, but they do not guarantee a positive experience.

The CEGE are based on the idea that a positive experience playing a game comes from the player's perception of the game (*video game* construct) and his interaction with it (*puppetry* construct) [28, 29]. The constructs of enjoyment and frustration were added [28, 29] to the Core Elements of Gaming Experience Questionnaire (CEGEQ) to create a total of four scales.

1. Enjoyment
2. Frustration
3. Puppetry
4. Video games

The CEGEQ is the instrument to assess the presence of these elements. The CEGEQ consists of 38 items that the participant must score on a 7-point Likert scale.

B. Interview structure

The interview was structured in correspondence with the levels of Kirkpatrick's model for training evaluation [106] (see Figure 6.1 on p. 122). The questions focused on the three bottom levels of the model: (1) reaction, (2) learning and (3) behaviour. Eight interviews were conducted according to a semi-structured approach, allowing room to elaborate on participants' specific answers and statements. The interview structure can be found in Appendix D.5.

All eight interviews were recorded using a smartphone, and handwritten notes were taken. For each interview, an informal report was written by the interviewer. The recordings were not transcribed verbatim.

6.3 Results

In this section, we will present the outcomes of the four parts of the study. We look at the quantitative and qualitative data from the questionnaires and the outcomes of the eight interviews. We also take into consideration the researcher's impression of the participants during the meetings.

We will first report on the Start sessions in Subsection 6.3.1. Next, we will report the outcomes of the three game blocks in Subsections 6.3.2, 6.3.3 and 6.3.4, respectively. Then, we will report on the repeating questions from all four questionnaires in Subsection 6.3.5. Finally, we will report on the eight interviews in Subsection 6.3.6.

6.3.1 Start session results

To start with the experiment, all participants visited the offices of NLR in Amsterdam. After a short introduction, the participants played the first game block of the *Shuttle to Mars* game, consisting of the tutorial and the first mission. After this, they answered the StM Start questionnaire.

Playing the first game block

All participants appeared to be enthusiastic about participating in the experiment and playing the game. They were interested in the research and the game.

Six participants played the game on laptop computers, two on desktop computers. On the laptop computers, the game was controlled using the touchpad or the keyboard. Three participants had difficulties operating the game via the touchpad.

While playing the tutorial, four participants requested additional information or explanation from the researcher. The researcher answered their questions to make sure they were able to play the game.

Five participants played the game quietly, while the other three were more outspoken. They expressed their engagement in the game with outcries or comments.

Initial questions

The initial questions were related to the participants' attitude towards innovations and their motivation to participate. All participants felt positive about the use of innovative training methods for airline pilots. Also, they were all positive about the use of virtual training or a serious game for training. Two participants, an instructor and a pilot, indicated that they were not usually pioneers in the use of new technology or software. The others were more supportive of the positive statement towards innovation.

The participants gave an average of 8.5 out of 10 ($SD = 1.1$) for their motivation to participate in this experiment. Except for one participant who gave a score of 3, all participants looked forward to playing the game.

With regard to playing the first part of the game, the participants were not convinced that the tutorial and the first mission gave a sufficient explanation of the game and provided sufficient practice ($M = 4.0$, $SD = 1.2$).

ICAO core competencies

Part of the StM Start questionnaire was related to the ICAO core competencies (see Appendix D.1). The participants were asked to score the importance of each of the eight ICAO core competencies in normal and non-normal situations as well as in emergencies. The competencies were scored on a 7-point Likert scale, ranging from *not at all important* to *very important*.

We observed differences between the opinions of the instructors and those of the pilots. Therefore, we show them separately in Figure 6.3. Due to the small number of instructors and pilots involved, the results must be interpreted carefully.

Overall, the pilots in our study believed that all competencies become more important in non-normal situations and emergencies. The instructors indicated that the importance of COM, L&T, SA and WM remains the same in all situations.

Interestingly, the pilots indicated that they believe AP becomes more important in emergencies, while instructors indicated a decrease in importance. A second interesting difference can be seen in the importance of COM. The instructors in our study attributed a higher importance to COM than the pilots did.

To complete the StM Start questionnaire, the participants were asked which training delivery method they believed to be most suitable for each competency (see Table 6.2). A description of training delivery methods is available in Appendix A.1. The ICAO core competencies are presented in Appendix C.2.

Pilots, as well as instructors, preferred the full flight simulator for training the competencies AFPM-A, AFPM-M, PS&DM, and SA. Also, the full flight simulator and the pc-based simulator were mentioned for the training of AP and WM competencies. For the training of COM and L&T, other methods than simulators were preferred. Serious games were mentioned in relation to the training of AP, PS&DM, SA, and WM.

6.3.2 Block 1 results

In Block 1, each participant played the set of five missions and answered the StM Block 1 questionnaire. All participants finished Block 1 on schedule. Two participants contacted the researcher during Block 1, enquiring after the number of missions in Block 1, the game controls and the need to follow the designated route.

The participants spent between 1 and 2 hours playing the *Shuttle to Mars* game ($M = 1.38$, $SD = 0.4$), according to their estimates in the StM Block 1 questionnaires. Two participants played other games during the Block 1 period.

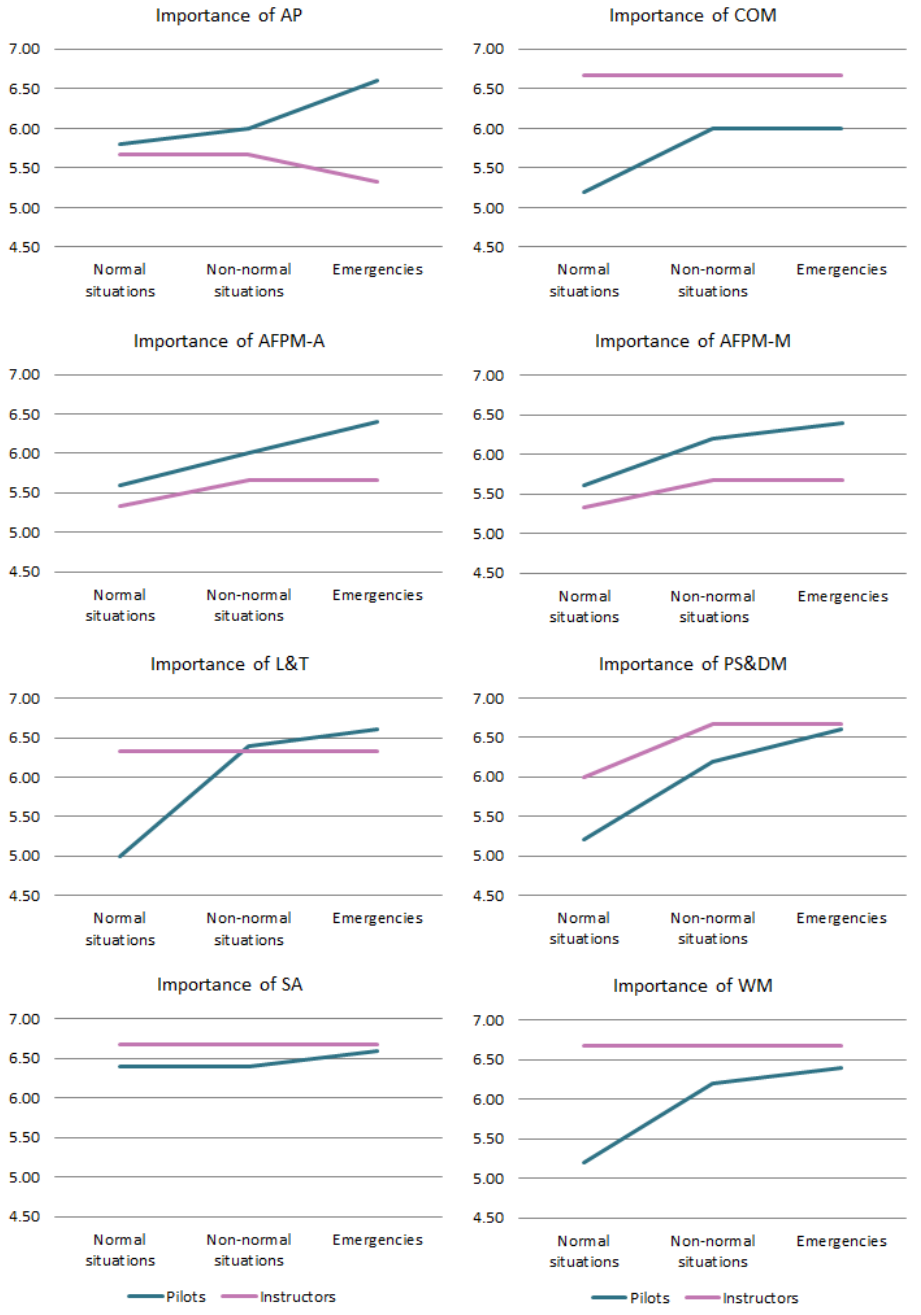

Figure 6.3: Importance of ICAO core competencies: comparing instructors with pilots

Table 6.2: Votes for most suitable training method per ICAO core competency

Training method	ICAO core competencies															
	AP		COM		AFPM-A		AFPM-M		L&T		PS&DM		SA		WM	
	P	I	P	I	P	I	P	I	P	I	P	I	P	I	P	I
Book/syllabus	2								1		1					
CBT		1	2													
Full flight simulator	2	1			4	3	5	3			2	2	3	3	1	1
Lecture/presentation				1						1		1				
PC based simulator		1	1		1										1	
Serious game	1										1		2		3	2
Training with a coach			2	2					4	2	1					

Note. P indicates the pilots' votes (n=5), and I indicates the instructors' votes (n =3).

User Experience Questionnaire

The StM Block 1 questionnaire comprised the UEQ [164, 181]. *Shuttle to Mars* received neutral scores for all scales, see Figure 6.4. The blue blocks represent the mean scores. They are almost zero on all six scales. These values are within the yellow area, indicating a neutral evaluation. Due to the small number of participants and the relatively wide range of scores, the consistency of the scales is rather low. Therefore, the scales should be interpreted very carefully [181].

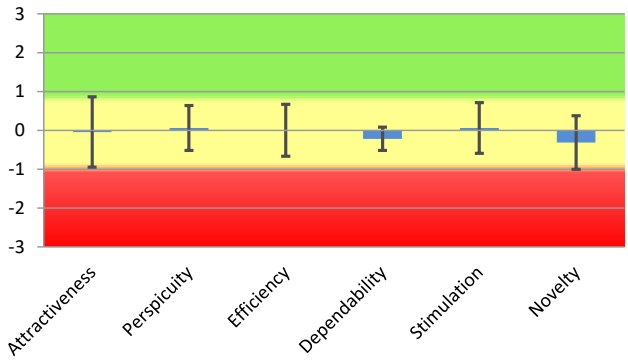

Figure 6.4: Shuttle to Mars' average UEQ scores per subscale with error bars

Using the UEQ benchmark, *Shuttle to Mars* can be compared to other interactive products. Products are classified into five categories per scale: (1) excellent, (2) good, (3) above average, (4) below average, and (5) bad. Based on the results, *Shuttle to Mars* falls into the lowest category *bad*, which contains the 25% worst results (Figure 6.5).

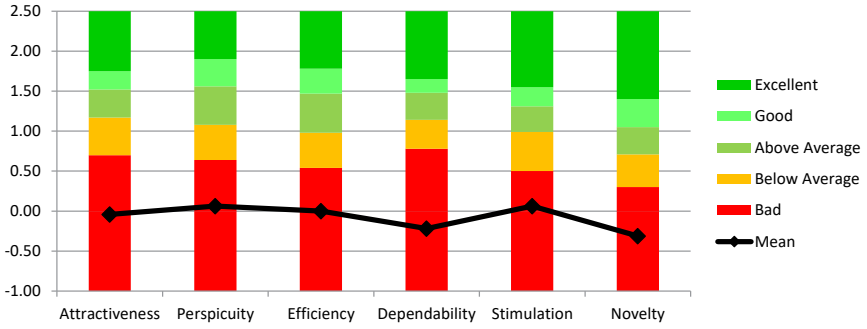

Figure 6.5: Shuttle to Mars comparison to UEQ benchmark

Reflective questions

The last part of the StM Block 1 questionnaire held reflective questions, to help the participants to relate the priorities in the game to the priorities in their daily jobs.

Participants could check a maximum of two priorities or enter their own. Six participants indicated that playing the game by the rules was a priority. Delivering the cargo to the destination was a priority for four participants. For only one participant, having fun was a priority.

Participants were asked to elaborate on their choice of priorities and to compare them to their priorities in their work. The participants recognised similarities but did also see differences. Below, we reproduce four quotes, taken from the questionnaires.

> *"Setting priorities and swiftly adjusting them when circumstances dictate to do so is recognisable from my work."*
>
> *"Somewhat comparable, but a lot of things that make my work easier are overly complicated in this game."*
>
> *"During my work, I am more focused on flying the airplane first, rather than solving a problem first or entering a code or answering a call."*
>
> *"The game challenges you to become better with every new stage of the game in an enjoyable way."*

Most of the participants seem to have used their skills as a pilot to play the game. They tried to reach the game objective while playing by the rules, and they tried to improve their game score.

6.3.3 Block 2 results

In Block 2, the participants played four missions and answered the StM Block 2 questionnaire. All participants finished Block 2 on schedule.

The participants spent between 1 and 2 hours playing the *Shuttle to Mars* game ($M = 1.6$, $SD = 0.3$), according to their estimates in the StM Block 2 questionnaires. Five participants played other games during the Block 2 period. Time spent playing other games ranged from 15 minutes to 6 hours.

Intrinsic Motivation Inventory

The StM Block 2 questionnaire comprised the IMI [91, 172]. Intrinsic motivation is considered to be measured with the Interest/Enjoyment subscale. On the Interest/Enjoyment subscale, participants scored a mean of 4.2 ($SD = 1.6$), indicating that their intrinsic motivation was not strong. Perceived Choice and Perceived Competence are thought to be positive predictors of intrinsic motivation, whereas Pressure/Tension is a negative predictor [91, 172]. The subscale of Perceived Choice was positive ($M = 5.42$, $SD = 0.6$), whereas Perceived Competence was neutral ($M = 4.25$, $SD = 1.2$). The subscale of Pressure/Tension was low ($M = 2.6$, $SD = 0.5$).

We added three items at the end of the instrument to enquire about the extent to which participants felt taken seriously and felt it was important to play the game seriously. Two participants indicated that they did not feel like they were taken seriously as professionals. The others were neutral or positive (n = 6, $M = 5.5$, $SD = 1.2$). All participants played the game seriously ($M = 5.4$, $SD = 0.7$). Also, we asked if they thought that making errors in the game *should* feel the same as making an error during a (simulator) training flight. Conspicuously, one participant fully agreed with this statement, one participant fully disagreed. The others were neutral or in slight agreement.

Reflective questions

The first two reflective questions in Block 2 were related to the participants' reasoning around the pirate attacks. Participants were asked on which considerations they based their actions and why. Five out of eight participants took their amount of resources into consideration, to decide whether to pay off the pirates. Also, five out of eight looked at the physical state of the spaceship (the *hull integrity*) to decide whether they would be able to survive an attack.

Reasons to engage the pirates included not wanting to pay cargo, and wanting to defend the shuttle before the pirates attacked. Two participants admitted that shooting the pirate was just fun. One participant avoided using the weapon because he did not understand how to use it. Reasons to pay the pirates included wanting to prevent damage, not wanting to pay a higher tax, and not wanting to go off track.

The other reflective questions were four questions enquiring about how the participants felt during the less eventful periods, and one open-answer question on how they deal with such periods in the cockpit. During the less eventful periods in the game, six participants became somewhat bored and did not pay as much attention during those periods. Four participants were somewhat distracted, and the other four were not. Five participants became a bit more watchful, as they expected that something would happen soon. For one participant, the less eventful moments did not change how he felt and played. The other participants indicated that these moments did have some influence.

Participants were asked to describe what they do in order to stay alert in the cockpit when not much is happening.

> *"I try to stay alert by talking with my co-worker. Now and then, I force myself to make a check around the flight instruments to see if everything is okay."*

Five participants indicated that they actively try to stay focused by checking their systems and by preparing for what lies ahead. Four participants said they talk to their colleagues to stay alert. The overlap is explained by one pilot who stated to do both.

6.3.4 Block 3 results

In Block 3, the participants played five missions and answered the StM Block 3 questionnaire. Again, all participants finished Block 3 on schedule.

The participants spent between 1 and 1.5 hours playing the *Shuttle to Mars* game ($M = 1.3$, $SD = 0.2$), according to their estimates in the StM Block 3 questionnaires. Two participants played other games during the Block 3 period, for 3 and 12 hours.

Three participants more or less enjoyed playing the game, but overall, the outcome was slightly negative ($M = 3.4$, $SD = 1.6$). For engagement, the scores were similar ($M = 3.3$, $SD = 1.7$). The three participants that enjoyed playing also indicated that they felt engaged in the game. Only one participant thought the assignments in the game were interesting. The others did not ($M = 2.1$, $SD = 1.4$). The participant who found the assignments interesting was also the only one who would recommend the game to others. Five participants could see the resemblances between the game and their job as an airline pilot. The other three participants gave a score of 2, indicating they do not see the resemblances.

Core Elements of Gaming Experience Questionnaire

The StM Block 3 questionnaire comprised the CEGEQ. Participants were slightly negative about their enjoyment of the *Shuttle to Mars* game. They gave it an average score of 3.3 ($SD = 1.6$) on the enjoyment scale. Only two participants gave a positive score. Participants gave an average score of 3.4 ($SD = 1.5$) on the frustration scale, indicating they were not really frustrated by the game. One participant indicated to have been frustrated by giving a score of 6. The subscales of the video game construct and the puppetry construct were all neutral, with scores between 3.9 and 4.4 out of 7.

Reflective questions

In the StM Block 3 questionnaire, one reflective question asked the participants to connect the events that happened in the *Shuttle to Mars* game to airline pilot tasks. The entry of authorisation codes, following the designated path, and the passing of other spaceships were easily connected to airline pilot tasks. Most participants answered in similar ways for these events. Participants had more difficulties relating the pirate attacks and negotiations with their everyday job.

The other two reflective questions enquired whether the participants believe they could learn something from the game and whether it could help them do their job in the cockpit. They were asked to score these statements on a 7-point scale and then elaborate on that score.

Overall, the participants gave a neutral score ($M = 3.9$, $SD = 1.0$) on being able to learn from the game. Two participants were a bit negative, four were neutral and two slightly positive. Learning to prioritise, stay focused, follow procedures and continuously scan the systems are the learning objectives most mentioned. Below, we reproduce two quotes, taken from the StM Block 2 questionnaire.

> *"In this game, you need to decide which tasks have priority and which do not. Also, you need a structure to scan all the instruments, while performing other tasks."*

> *"To be able to learn, sometimes you need feedback and tips while you are doing it or after you have done your task."*

On the question of whether the game could help them do their job, participants were somewhat more negative ($M = 3.3$, $SD = 1.4$). Four participants gave a negative score, two were neutral and two slightly positive. Three participants stated that the game does not resemble the actual work environment sufficiently to be helpful. Others stated that they believe it could help them in prioritising, problem-solving and multitasking. Also, it was mentioned that the game could be useful for memory items. Memory items are essential checklists that a pilot should not need to look up, but that should be committed to memory.

6.3.5 Repeating questions

Several questions in the questionnaire were asked multiple times. Three questions from the StM Start questionnaire were repeated in the StM Block 3 questionnaire. Furthermore, five questions were part of all questionnaires.

Questions asked after Start session and Block 3

Three questions were asked at the end of the StM Start questionnaire, and again in the StM Block 3 questionnaire. These questions were related to (1) Purpose, (2) Control, and (3) Graphics. Results for these repeating questions are visualised in Figure 6.6.

The first question asked whether the purpose of the game was clear to the participant. Throughout playing the game, the purpose of the game became less clear to the participants. After the Start session, only one participant gave a slightly negative score. On average, the participants were positive about how clear the purpose of the game was ($M = 5.3$, $SD = 1.0$). After Block 3, two participants indicated that the purpose of the game was not clear to them. The others were more positive, resulting in a neutral overall outcome ($M = 4.4$, $SD = 1.3$).

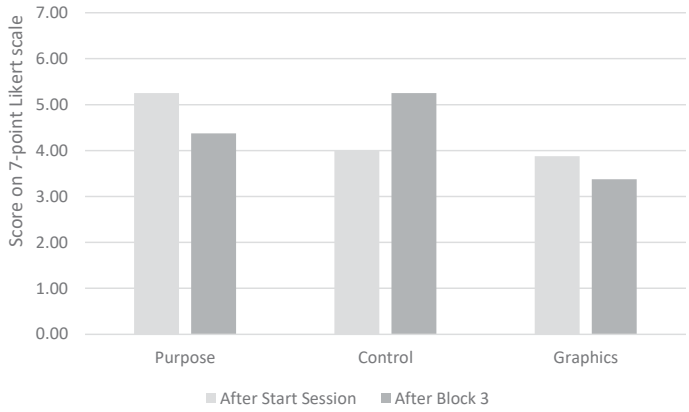

Figure 6.6: After Start session and StM Block 3: Purpose, Control, Graphics

The second question enquired about how much the participant felt he was in control of the game. Playing the game resulted in the participants feeling to be more in control over the game. The feeling of being in control improved in the period between the Start session ($M = 4.0$, $SD = 1.1$) and the end of the study after StM Block 3 ($M = 5.3$, $SD = 0.9$).

The third question asked how much the participant liked the graphics of the game. The low average scores showed that the participants did not really like the graphics of the game. At the end of the study, the appreciation of the graphics decreased. After the Start session, participants gave an average score of 3.9 ($SD = 1.5$). One participant did not like the graphics at all. He gave a score of 1. After Block 3, one participant indicated to like the graphics, but all others did not. This resulted in an average score of 3.3 ($SD = 1.5$).

Questions asked in all four questionnaires

Five questions were asked in every questionnaire. The first question asked how the participant felt after completing part of the game (Figure 6.7). Towards the end of the study, more participants were relieved to have completed an StM Block.

After the Start session, five participants expressed that they had wanted to continue playing, two were neutral, and one was relieved. After Block 1, three participants expressed that they had wanted to keep playing. Three were neutral, and two were relieved. After Block 2, four participants expressed that they had wanted to keep playing. Two were neutral, and two were relieved. After Block 3 was completed, only one participant expressed that he had wanted to keep playing. Two were neutral, and five were relieved.

The other four questions asked the participants to give a score from 1 to 7 on four different statements (see Figure 6.8).

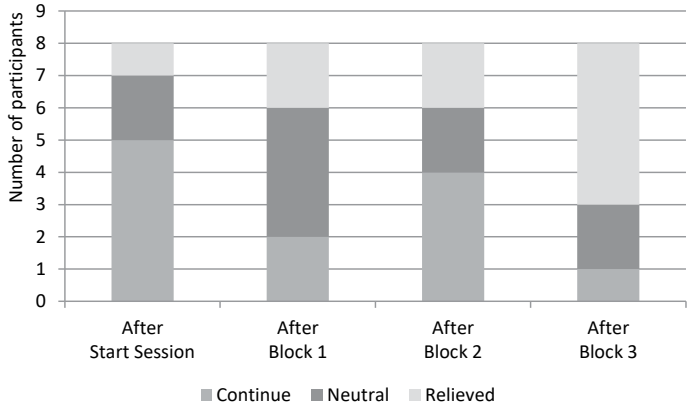

Figure 6.7: After each StM Block: Players feelings about completing the StM block

The first question of the remaining four (labelled as *Motivation*) showed that on average the participants were more motivated to complete the experiment than to play the game. Only after Block 2, two participants indicated this statement was not true, giving it a score of 3 out of 7.

The second question (labelled as *Look forward*) asked the participants how much they looked forward to playing the next mission. At first, participants looked forward to the next mission, but in the end, this anticipation decreased.

The third question (labelled as *Frustrated*) enquired about the extent to which playing the game had frustrated the participants. Overall, the participants were not frustrated by the game, but there was a slight increase in frustration towards the end. One participant who did not look forward to the next mission scored a 5 on having been frustrated by playing the game after the Start session. After Block 3, three participants indicated that they had become frustrated by the game.

The fourth question (labelled as *Commercial*) showed that the participants never thought the game could compete with commercial games. Playing the game did not change their opinion much. They became slightly more negative after playing Block 1 and then remained constant.

6.3.6 Interviews

We used a semi-structured interview to debrief the participants about their experience with the *Shuttle to Mars* game (see Appendix D.5). The interview was structured into three categories, that are related to the three lower levels of Kirkpatrick's model of training evaluation [106], viz. (1) reaction, (2) learning, and (3) behaviour.

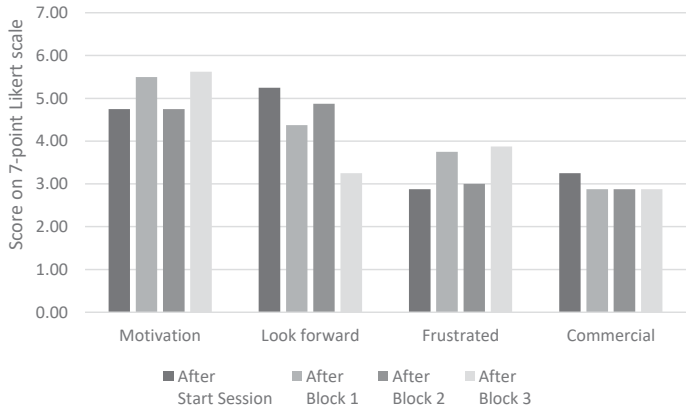

Figure 6.8: After each StM Block: (1) Motivation, (2) Look forward, (3) Frustrated, (4) Commercial

Question category: Reaction

The questions in the Reaction category aimed to gauge the participant's feelings about the game and about having to play the game.

Only one participant did not like the game, neither as a game nor as a training method. This participant classified himself as an experienced gamer. Four participants said that later on in the game, they became bored or a bit annoyed with the game. In contrast, one participant did not like the game until later on, when he got the hang of it.

All participants could imagine the game being part of a training, but for most, it would have to be a module early on during initial training. Several participants suggested the game could be used as a tool for candidate selection for flight schools.

Most participants believed this game could not help them in their jobs. This was mostly because they thought the game would be more beneficial to less experienced pilots.

Several participants said that the Beta 2 version of the *Shuttle to Mars* game did not have sufficient quality. The visuals looked dated, and the gameplay was not sufficiently challenging. Some participants were bothered by the music. The main thing the game was lacking, according to participants, was feedback on their actions. They could not see how their score was influenced by what they did. Therefore, it was impossible for them to change their behaviour to improve their score. Some participants also indicated that the game should resemble an actual aircraft a bit more.

Question category: Learning

The questions in the Learning category asked about what the participants have learned and how they would compare the competencies between the game and their actual job.

Participants believed that novice pilots could learn about prioritising, working under stress, the importance of adhering to procedures and the importance of staying alert. However, they indicated that they themselves did not really learn anything from the game, more than how to play the game.

In the interviews, the participants indicated that they believed the competencies in the game were comparable to those in real life, although they were not identical. All participants recognised some of the ICAO competencies they used to play the game. The competencies named by the participants coincided with three of the competencies the game aimed to train, i.e., PS&DM, SA, and WM.

The participants believed that the competencies could be trained with a game, especially when dealing with less experienced pilots, but not with the *Shuttle to Mars* game in its current form. Hence, they did not believe their competencies had become stronger by playing the game.

Question category: Behaviour

The questions in the Behaviour category focused on possible effects of the game on their behaviour, especially their behaviour in the cockpit. Such effects could be considered an early sign of transfer [212].

Six participants did not spend time thinking about the game, besides playing the game or planning when to play the game. One participant showed a co-worker a video that he had made of the gameplay. They discussed the below-standard quality of the game. Another participant was reminded of the game when he was in a situation in which three things happened at once in the cockpit.

All participants believed that the *Shuttle to Mars* game could have a positive effect on novice pilots. However, they did also believe the game had not influenced how they themselves operate in the cockpit. Playing the game longer and more often may have more effect, but most participants also said that the game needed to be improved.

6.4 Discussion

In summary, we conducted the following investigations. During a four-week period, eight airline pilots participated voluntarily in the *Shuttle to Mars* study. They came to the offices of NLR for a Start session, then played the game for three weeks. After they completed the game, they returned to NLR for the final interview.

In this section, we will discuss the outcomes of our study. We will look at the participants' motivation (Subsection 6.4.1), the game experience (Subsection 6.4.2), and games as a training method (Subsection 6.4.3). Finally, we will briefly discuss the quality of the *Shuttle to Mars* game (Subsection 6.4.4).

6.4.1 Motivation

We look at motivation to determine whether airlines pilots would be willing to play the serious game and to keep playing it over an extended period.

The participants were motivated to participate in the experiment. From the questionnaires, we saw that their motivation was more about completing the experiment than about completing the game itself. This attitude remained stable throughout the experiment, from the start until Block 3 (Figure 6.8). Although several participants indicated that after some blocks of gameplay, they were "done with the game", they remained motivated to continue, and all participants completed the experiment.

Overall, the participants were positive about innovative training methods and the use of virtual training or a serious game for training. This may have had a positive effect on their motivation to participate in this study.

The IMI data [91] indicated that only part of the participants' motivation is intrinsic. The participants also had an extrinsic source of motivation. This may have been the chance of winning the smartwatch, or it may have been the mere fact of participating in an aviation-related study. However, they did feel that they had a choice and were not pressured to participate and to play the game.

In a training setting, especially during initial training, aspiring pilots will be motivated to do whatever is needed to get their pilot licence. This would also contribute to the motivation to play the game.

Participants in the study showed a slight decrease in motivation over time but still appeared enthusiastic and motivated in the final interviews. This may indicate that, with improvements in the game, the game may become an enjoyable and thus motivational part of training.

6.4.2 Game experience

At the beginning of the study, the were enthusiastic and motivated. During the interview at the end, they appeared still to be enthusiastic and motivated. However, from the data, we may conclude that playing the *Shuttle to Mars* game was not a truly positive experience.

Although the UEQ data [164, 181] was neutral, we may conclude that the *Shuttle to Mars* is not an adequate interactive product, which is confirmed by the CEGEQ data [28, 29]. The core elements of having a sufficient game experience are not convincingly present in *Shuttle to Mars*.

One of the repeating questions asked was whether the gameplay of the *Shuttle to Mars* game could compete with that of commercial games. During the experiment, this value was consistent (Figure 6.8). The score is slightly below neutral, indicating that the participants believed that the gameplay of the *Shuttle to Mars* game could not compete with commercial games. During the interviews, the participants made suggestions for improvements to the game.

In contrast, other data from the questionnaires indicated that playing the game was not an unpleasant experience either. With improvements to the game, the experience may become enjoyable. The UEQ and CEGEQ data may serve as guidance by which aspects of the game may be improved. Participants also made several suggestions on how to improve the game.

The suggestions for improvements to the game will be discussed in Subsection 6.4.4.

6.4.3 Games as a training method for competencies

Interestingly, most participants reported that they did not learn anything from the game, but they did believe that other, less experienced pilots may benefit from it. Participants acknowledged that the competencies of PS&DM, SA, and WM were addressed by the game, as well as AP. They were able to relate game events and game tasks to specific tasks in their jobs. The participants believed the game may best be used by aspiring pilots during initial training. The Beta 2 version of the *Shuttle to Mars* game was not ready to be used as a training method, but the participants did see the potential.

Some participants would like the game to resemble an actual aircraft. The game was intentionally designed as a zero-fidelity simulation [192, 193] to focus on the competencies instead of the environment. Using a game environment that resembles an actual aircraft may put the focus on the type of aircraft and the way in which the game environment deviates from reality.

6.4.4 Quality of the Shuttle to Mars game

Although the study was not aimed at evaluating the quality of the *Shuttle to Mars* game or its effectiveness, we did receive feedback on this. As many participants remarked, the quality of the game was insufficient. Before a final version can be implemented in the airline pilot training curriculum, improvements will have to be made, to make playing the game an enjoyable experience. In the questionnaires and during the interviews, several suggestions were made on how to improve the game (see Section 5.5.2). In addition to those suggestions, participants would like to see the game resemble current commercial games in visualisations and gameplay.

Suggestions from the small-scale study participants

The Beta 2 version that was used in the *Shuttle to Mars* study consisted of fifteen missions. Participants played on personal computers. There were no technical issues hampering gameplay. Most of the suggestions made by the participants in the *Shuttle to Mars* study are related to functions that were omitted in the beta versions or game content.

Game control. As the game was designed for touchscreen interaction, controlling it with a touchpad or mouse was not optimal. If the final version is going to be played on personal computers, the control needs to be improved for keyboard and mouse control.

Tutorial. The tutorial does not provide sufficient support to play the game well. Some actions remain unclear and should be addressed better in the tutorial. Moreover, the current tutorial raises expectations that the game does not meet. The tutorial should reflect the actual gameplay. It should not cover more than is needed. Possibly the tutorial could be replaced with a number of training missions.

Feedback. The game should provide feedback on what the player has done. This could be at the end of a section, a sector or a mission. It should address the behaviour that was (un)desirable, or that has influenced the score.

Score. It should be clear why a specific score was given. The player should know whether his choices or behaviour caused his score to be lower. This may take place either in advance or right after the first incident, so that he can improve his score.

Redundant tasks. Currently, the game has some redundant tasks. When playing, the player will come across them, but they do not make a difference in the game. For instance, each mission must be planned, but there is only one possible route in the beta versions. At the end of a sector, resources can be bought and sold in order to continue the mission, but this is never necessary during the Beta 2 version of the game. These tasks should either be implemented or removed from the game.

Missions. The game should have a larger number of missions. Also, the missions need to be more complex and more challenging. The missions should be matched to situations that may actually happen in the cockpit. To achieve this, close cooperation with flight instructors is needed in designing the missions.

After the *Shuttle to Mars* study, no further developments have been done on the Beta 2 version. The improvements described here have not been applied.

When the development of the *Shuttle to Mars* game is resumed, we advise that the final version of the game be developed for use on iPads. Controlling the game through mouse or keyboards is difficult, and the playability of the game will benefit from touchscreen interaction. Elements in the game environment that do not need to be used in the game tasks should be removed, or their presence should enhance the narrative.

Also, from an instructional design point-of-view, there are improvements to be made. Although the game was designed according to our SG4CD model, not all elements were sufficiently developed. Improvements can be made in two areas specifically, viz. (1) support and feedback, and (2) the match between the game and aviation. First, the game should provide the player with more support and feedback. This will help the player to feel in control over the game, and it will also support the learning effect. Giving more (cognitive) feedback during the game, will also help the player strengthen his competencies. Second, closer cooperation between flight instructors and aviation training professionals on the design of the meaningful events may result in a stronger transfer from game to job.

6.5 Chapter conclusion

In this section, we will draw our conclusions about the findings from our study. As we have discussed in the introduction of this thesis, younger pilots nowadays lack certain competencies. The purpose of this study was to see whether airline pilots would accept serious games as a training method to reduce this deficit.

We will answer the third research question in Subsection 6.5.1. Then, we will give our recommendations for future research in Subsection 6.5.2.

6.5.1 Answering research question 3 (RQ 3)

The *Shuttle to Mars* study was performed to answer research question 3.

RQ 3: *To what extent do airline pilots accept a game to develop essential competencies for critical situations?*

From the results of our research, we may conclude that airline pilots are open to the innovative approach of using a game to train the essential competencies. Playing the current version of the *Shuttle to Mars* game did neither provide a convincingly positive experience, nor was the experience wholly negative. With improvements to the game, and embedding it in the initial training for airline pilots, the game has the potential of providing a positive and effective experience.

In terms of Kirkpatrick's model for training evaluation [105, 106], the outcomes of our study constitute a positive *reaction* (level 1). This positive reaction is ground for optimism about the use of serious gaming for the training of airline pilots.

We started out with the *Shuttle to Mars* game, aiming it to be used in the competency development of young, but graduated pilots. As a result of the study, we now think that aspiring pilots may benefit from the game during their initial training. Making sure that they develop these competencies during flight school, may give them a profound basis to start from as a professional. Still, we do believe that an improved version of the *Shuttle to Mars* game embedded in a learning package with sufficient feedback can also be useful for experienced pilots.

Applying the game in initial training may also move ahead the training of more complex scenarios and prepare the aspiring pilots for training sessions in the full flight simulator. This may provide some relief on the simulators, as later on in the programme, simulator sessions can be dedicated to technical skills.

Therefore, our provisional answer to RQ 3 is that airline pilots have a positive attitude towards the idea of game-based learning to develop competencies and are willing to accept it as a training method.

6.5.2 Future research

Before further research can take place, improvements to the *Shuttle to Mars* game are essential. Further research should then focus on (1) the learning effect of the game, (2) the transfer of competencies from the game to the cockpit, and perhaps (3) exploring possible effective applications for the *Shuttle to Mars* game. This may include selection of candidate pilots, training candidate or novice pilots (acquiring competencies), or training experienced pilots (maintaining/boosting competencies).

Measuring the learning effect of the game

The learning effect of the game can be determined by analysing the in-game data. Players are expected to improve their performance in the game over time. Ideally, this improvement can be attributed to the development of the competencies. Especially the

performance in new situations may provide indications of the learning effect. Either these new situations should be similar to earlier situations, or they should address the same competencies as earlier situations under different circumstances.

Measuring the transfer of competencies

Once a learning effect of the game has been found, ideally it should transfer to the work environment. This transfer may be proven with a random controlled trial, using a flight simulator to assess participants' competencies in the work environment, before and after playing the game.

Using Shuttle to Mars as a selection tool

During our study, participants suggested the *Shuttle to Mars* game may be useful as a selection tool for flight school candidates. The performance of candidates in the game may be an indication of their potential performance in the cockpit, and their ability to become a good pilot. To test this use of the game, the game should be compared to selection tools that are currently being used to find qualified candidates.