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

Kuindersma, E.C.

Citation

Version:	Publisher's Version
License:	[Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)
Downloaded from:	[https://hdl.handle.net/1887/77168](#)

Cover Page

Universiteit Leiden

Author: Kuindersma, E.C.
Title: Cleared for take-off: Game-based learning to prepare airline pilots for critical situations
Issue Date: 2019-10-24

Chapter 4

The CloudAtlas game:
Voluntary play in serious games

In this chapter, we will address RQ 2, which reads as follows.

RQ 2: *What is the effect of voluntary play on the outcomes of a serious game?*

Voluntariness is an important feature of games. Several scholars in the field of games list voluntariness as one of the main characteristics [26, 69, 89, 133, 135] (see also Section 2.5.3).

To the best of our knowledge, no studies have taken into account the possible effect of voluntariness on learning and gameplay within game-based learning (i.e., voluntary versus mandatory gameplay). To fill this hiatus, we performed a series of three experiments. We aim to determine to what extent the learning effect and the gameplay experience of a serious game are affected by the student's freedom to choose to play the game.

This chapter is organised into seven sections. In Section 4.1, we will give an overview of how we set about to measure the effect of voluntariness by conducting three experiments. The three experiments are then discussed one after the other in three subsequent sections: Sections 4.2, 4.3, and 4.4. Then, in Section 4.5, we will discuss the three experiments as a whole. Section 4.6 describes the limitations we faced. Finally, in Section 4.7, we will answer the research question.

This chapter is based on two previous publications:

1. Kuindersma, E. C., van der Pal, J., van den Herik, H. J., & Plaat, A. (2015). Voluntary Play in Serious Games. In International Conference on Games and Learning Alliance (pp. 131-140). Springer International Publishing.
2. Kuindersma, E. C., van der Pal, J., van den Herik, H. J., & Plaat, A. (2016). Comparing Voluntary and Mandatory Gameplay. International Journal of Serious Games, 3(3), pp. 67-83.

4.1　Measuring the effect of voluntariness

With our series of three experiments, we aim to measure the effect of voluntariness on the outcomes of playing serious games. In Subsection 4.1.1, we will first identify the outcomes of playing serious games. Then, in Subsection 4.1.2, we will describe our expectations about the effect of voluntariness. In Subsection 4.1.3, we will discuss the general set-up of the experiments. A comparison of the three experiments will be provided in Subsection 4.1.4.

4.1.1　The outcomes of serious games

In this study, we examine the effect of voluntariness on the outcomes of a serious game. We distinguish two types of outcomes, viz. (A) the learning effect of the game, and (B) the gameplay experienced by the player. Below, we discuss and define both types of outcomes.

A. Learning effect

The learning effect is the main type of outcome in an investigation of serious games. The primary goal of a serious game is to make the player learn something (see Definition 2.2). Each game has specific *learning objectives* (Definition 4.1) that describe what the player will learn from the game. After playing the serious game, the player will demonstrate to have achieved a *learning outcome* (see Definition 4.2), e.g., by demonstrating specific behaviour or successfully taking a test. In an ideal situation, the learning outcome will match the learning objective.

This study focuses on the *learning effect* (Definition 4.3) of a serious game. The learning effect is not necessarily identical to the learning outcome. Part of the learning outcome may be a result of something other than playing the serious game, e.g., prior knowledge, the use of other learning materials, or the interaction with other players. It may even be a result of an unrelated activity. An *effective training method* (Definition 4.4) is successful in producing the desired result [57], i.e., in achieving the learning objective.

Definition 4.1 - **Learning objective**

A learning objective is a statement that defines the intended goals of a learning activity in terms of the knowledge and skills that the learner will acquire as a result of the learning activity. It describes an intended *state.*

Definition 4.2 - **Learning outcome**

A learning outcome is a statement that describes the knowledge and skills a learner has achieved and demonstrated upon completing the learning activity. It describes an observed *state.*

Definition 4.3 - Learning effect

The learning effect is the part of the learning outcome that can be attributed to the learning activity.

Definition 4.4 - Effective training method

A training method is effective when it produces the desired result, i.e., the intended learning effect.

B. Gameplay experience

The second type of outcome of playing a serious game is how playing the game is experienced by the player. In our studies, we look at the gameplay experience as a combination of the player's motivation, his enjoyment, and engagement. Below, we will define motivation, enjoyment and engagement.

In a setting of mandatory play, players will be obliged to play. They are not free to choose whether to play or not (i.e., play is an imposition). This may have a negative effect on the game experience [84, 138]. Therefore, we also take into account a player's feeling of being subject to an obligation.

Definition 4.5 - Motivation

Motivation is the willingness to participate, and the enthusiasm and determination with which a player participates.

Definition 4.6 - Enjoyment

Enjoyment is the extent to which the player takes pleasure or satisfaction in participating.

Definition 4.7 - Engagement

Engagement is the extent to which the player is involved or committed to his participation.

The meanings of motivation, enjoyment, and engagement are overlapping, but the concepts are certainly not equal. A player can be motivated to play, yet not enjoy the game or never become engaged in the game. Similarly, a player may become engaged in a game he was not motivated to play. He may even be engaged in a game he believes he does not enjoy.

Based on the two types of outcomes of serious games, we split RQ 2 into two subquestions.

RQ 2a: *To what extent does the voluntary play of a serious game affect the learning effect?*

RQ 2b: *To what extent does the voluntary play of a serious game affect the gameplay experience of the player?*

To answer RQ 2 and its subquestions, we conducted an exploratory study to determine whether using a serious game as a learning tool voluntarily as opposed to mandatorily, has an effect on the outcomes.

The exploratory nature of the study guided us in our structuring of the study. We decided to have a series of experiments, in which the next experiment took into account the results of the previous experiments in its structure and design. We decided in advance to limit the number of experiments to three.

After each experiment, we will try to answer RQ 2a and RQ 2b with respect to the experimental conditions. Subsequently, we will answer RQ 2, using the outcomes of all three experiments.

4.1.2   Expectations

In conducting the series of experiments, we have six expectations. The expectations are based on literature about the effectiveness of serious games (see Subsection 2.5.2) and voluntariness (see Subsection 2.5.3). References are provided in the respective subsections.

1. **The game.** First of all, the CloudAtlas game that was designed for these studies is expected to be an effective learning game. We expect that players gain knowledge about clouds by playing the game, and that they are able to recognise the clouds and decide the best way to act around these clouds.
2. **Test score.** Furthermore, we expect a positive effect of voluntary play on the learning effect. We expect voluntary players to achieve a greater improvement of their knowledge of and insight into clouds than mandatory players. Voluntary players should achieve higher test scores.
3. **Game score.** As the knowledge of and insight into clouds is part of the gameplay, we expect voluntary players to do better in the game than mandatory players and achieve higher game scores.
4. **Enjoyment.** Also, we expect a positive effect of voluntary play on the gameplay experience. We expect voluntary players to enjoy the game more than mandatory players.
5. **Time spent playing the game.** As a result of the greater enjoyment, we expect that voluntary players will play the game for a longer time than mandatory players.
6. **Obligation.** Volunteers participate out of their own choice. Hence, it is to be expected that voluntary participants are motivated to participate, more so than mandatory participants. We even expect a negative effect of mandatory participation. Mandatory participants will have a negative feeling about being obliged to play.

4.1.3 General set-up of the experiments

All three experiments have a similar set-up and look for the same measurements. The research design of the three experiments is a combination of exploratory and experimental research. The participants study written materials on cloud identification, play a game (voluntarily or mandatorily), and then take a test. All participants are free to choose how much time they spend on studying the written materials.

Participants who play the game voluntarily are free to choose how long they play, including not playing at all. Participants who play the game mandatorily have to play for at least ten minutes. The procedure is described in more detail in Subsection 4.2.3, and the materials are described in Subsection 4.2.4.

The second and third experiment build on the previous experiments, using the same procedure and materials but with some adaptations. The adaptations made for the experiments are described in Sections 4.3 and 4.4.

In all three experiments, we differentiate between the participants based on two independent variables, viz. (1) participation, and (2) gameplay.

1. **Participation.** Participants who have volunteered to participate in the experiment are voluntary participants (VP). In contrast, the participants who partake in the experiments as part of a school assignment are mandatory participants (MP).
2. **Gameplay.** Participants who are free to choose whether they play the game and how long they play the game, are voluntary players (VG). In contrast, the participants for whom a minimum of ten minutes of gameplay is enforced, are mandatory players (MG).

Experiment 1

The first experiment (Figure 4.1) had an informal, non-educational setting. It took place in July and August 2015. A total of 19 voluntary participants (VP), with no link to aviation, were randomly assigned to either voluntary gameplay (VG) or mandatory gameplay (MG). Moreover, a reward was offered to the participants. The experiment yielded interesting outcomes that led to a second experiment. Experiment participation, even when assigned to the mandatory gameplay (MG) condition, has a much stronger voluntary character than GBL usually will have. Therefore, the second experiment was set up as part of a course to create some level of mandatory participation (MP). We will discuss the results of Experiment 1 in Section 4.2.

Experiment 2

The second experiment (Figure 4.2) was carried out from March to May 2016, with a total of 74 participants. A total of 71 participants had no link to aviation. The experimental design and materials were adapted to include mandatory participants (MP) in addition to the voluntary participants (VP), and to introduce a no-gameplay (NG) condition in addition to voluntary (VG) and mandatory (MG) gameplay. Some of the participants received an incentive or reward for participating. Again, the outcomes of the experiment
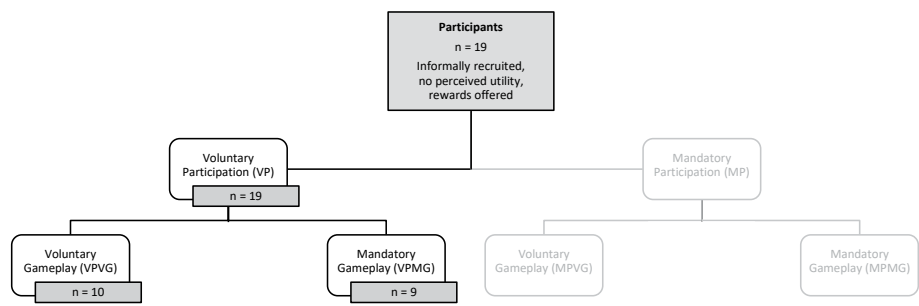
Figure 4.1: Structure and number of participants of Experiment 1

were interesting but inconclusive. Hence, a third experiment was conducted. We aimed to create an even stronger sense of mandatory participation (MP). We will discuss the results of Experiment 2 in Section 4.3.
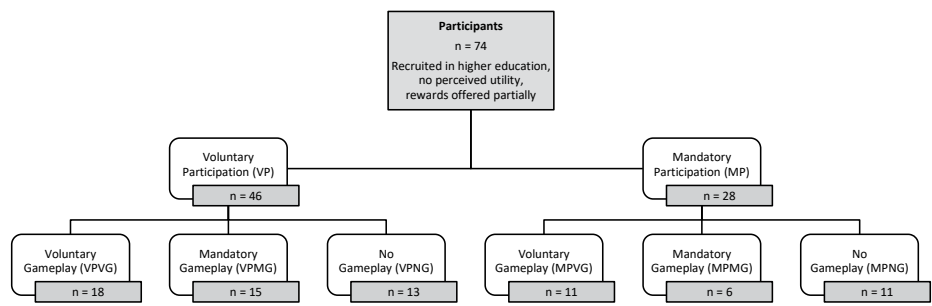

Figure 4.2: Structure and number of participants of Experiment 2

Experiment 3

The third experiment (Figure 4.3) took place between February and August 2017, with a total of 83 participants. This experiment had a formal setting with a stronger distinction between voluntary (VP) and mandatory (MP) participation. No incentives or rewards were offered to eliminate possible confounding variables. All participants were (aspiring) pilots. Therefore, the topic of the experiment had a stronger relevance and utility for them.

The third experiment confirmed, to some extent, the interesting outcomes of the previous experiments. We will discuss the results of Experiment 3 in Section 4.4.
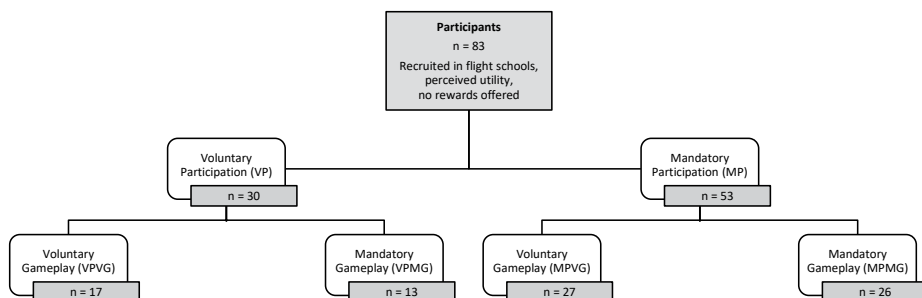
Figure 4.3: Structure and number of participants of Experiment 3

Overall discussion of the three experiments

In Section 4.5, we will give an extendsive overview of the results of all three experiments.

4.1.4 Comparing the experiments

Although the experiments have a similar set-up, there are differences between them in the areas of (A) independent variables, (B) setting, (C) incentives and rewards, (D) utility, and (E) the measurements for which we look. Below, we discuss these differences. An overview is given in Table 4.1.

A. Independent variables

In all three experiments, we compared playing a serious game voluntarily (VG) with playing the game mandatorily (MG).

In addition, in the second and third experiment, we looked at the effects of participating in the experiment voluntarily (VP) versus mandatorily (MP).

All participants in Experiment 1 participated voluntarily (VP). Hence, in Experiment 1, all participants are in the group of voluntary participation (VP). In Experiments 2 and 3, teachers assigned participation to their students, resulting in a group of mandatory participants (MP) in addition to the voluntary participants (VP).

The independent variables for all three experiments are (1) Participation and (2) Gameplay.

B. Setting

All participants in Experiment 1 were recruited through social media. They volunteered to participate. In Experiment 2, recruitment was done through institutions for Higher Education, and in Experiment 3 through Flight Academies.

Some participants volunteered; for others, it was part of a school assignment. This resulted in an informal setting for Experiment 1 and more formal settings for Experiments 2 and 3.

C. Incentives and rewards

In Experiment 1, all participants who completed the experiment had a chance of winning a gift card. This can be considered to be an incentive to participate. In Experiment 2, some of the VP participants had a chance to win a gift card, while other VP participants received school credits. Receiving credits can also be considered to be a reward. The MP participants were neither offered an incentive nor did they receive a reward. In Experiment 3, we did not offer any incentive or reward to the participants.

D. Utility

When utility is lacking, motivation to learn is unlikely [203]. Based on the work by Vroom [203] and Clark, Dobbins and Ladd [34], we define utility as follows.

Definition 4.8 - **Utility**

Utility is the perceived usefulness of a training to reach a particular goal, for example in a job or career.

Participants have a higher motivation to learn and report a greater amount of learning when they identify the utility of the training for their job or career [34, 168]. Furthermore, a higher sense of utility has been linked to a higher motivation to transfer the knowledge and skills to the workplace [121, 170].

For Experiments 1 and 2, we did not focus on recruiting participants for whom the topic of the experiment was relevant. A small number of participants may have perceived utility in the topic, but in general, the topic was not a motivating factor. Experiment 3 was aimed at pilots and pilots-in-training. Therefore, the topic was more relevant to them, i.e., they had a higher perceived utility.

E. Measurements

All three experiments in the study serve to determine the effect of voluntary (VG) and mandatory (MG) gameplay on the game's learning effect and the player's game experience. In total, we look at seven measurements (see below and Table 4.1) to determine the learning effect and game experience. Two measurements are objectively measured, viz. Game score and Test score, the other measurements are more subjective as the participants give their personal opinion about them. The participant's opinion about being obliged to play is used to measure the effect of mandatory play. After Experiment 1, we add a focus on the motivation to participate and the engagement in the game.

With regard to the learning effect (RQ 2a), we look specifically for Measurement 1: Game score, Measurement 2: Test score, and Measurement 3: Time spent playing the game.

With regard to the gameplay experience (RQ 2b), we look for Measurement 4: Enjoyment, and Measurement 5: Obligation. If present in the experiment, we will also look for Measurement 6: Motivation, and Measurement 7: Engagement.

Table 4.1: Comparing the experiments

	Experiment 1	Experiment 2	Experiment 3
Independent variables	Gameplay	Participation, Gameplay	Participation, Gameplay
Setting	Informal	Formal	Formal
Incentives and rewards	For all participants	For some participants	None
Utility	None	None	High
Measurements			
1. Game score	+	+	+
2. Test score	+	+	+
3. Time spent	+	+	+
4. Enjoyment	+	+	+
5. Obligation	+	+	+
6. Motivation	-	+	+
7. Engagement	-	-	+

4.2   Experiment 1: Informal Setting

The first experiment was set up to determine whether voluntarily using a game as a learning tool will result in a better performance on a test. It took place in an informal setting. The design of Experiment 1 is based on one independent variable, being Gameplay.

The design will be described first (Subsection 4.2.1), followed by the participants (Subsection 4.2.2), the procedure (Subsection 4.2.3), and the materials used (Subsection 4.2.4). Then, the results are presented (Subsection 4.2.5) and discussed (Subsection 4.2.6). Finally, conclusions on Experiment 1 are drawn (Subsection 4.2.7).

4.2.1   Design

The study employed a balanced mixture of quantitative and qualitative methods, using two groups for the independent variable of gameplay. Participants were randomly assigned to one of the two Gameplay[1] groups.

1. **The voluntary gameplay group (VG)** in which players were free to choose how long to play the game or not to play the game at all.
2. **The mandatory gameplay group (MG)** in which players had to actively play the serious game for a minimum of 10 minutes.

Remarks

1. In the first experiment, all participants volunteered. To allow a comparison with the second and third experiment, we will categorise all participants as voluntary participation (VP).
2. The independent variable was the type of Gameplay.

---

[1]Please note that in order to avoid confusion between participants and players, we use the abbreviation P for participation/participants and G for gameplay/players.

3. We focused on five measurements: (1) game score, (2) test score, (3) time spent playing the game, (4) enjoyment, and (5) obligation.
4. We did not take into account motivation and engagement.

4.2.2 Participants

Participants were recruited through social media (Facebook, LinkedIn, and Twitter) and by personal invitation. They were told that the experiment was related to aviation, but the focus on gaming was not disclosed. Only persons over the age of 18 were selected to participate. They were asked to give their informed consent before being registered. As an incentive, participants were offered a chance to win a €100 gift certificate. Chances of winning were related to completing all stages of the experiment, not to personal results.

A total of 64 persons registered for the experiment and completed the online pre-experiment questionnaire (Q1). The 64 participants were randomly assigned to one of two groups, resulting in a VG group of 29 participants and an MG group of 35 participants. Out of the 64 registered participants, 19 completed the experiment in a valid way, i.e., by studying the materials and taking the test. The other 45 participants failed to complete the experiment due to a variety of failures. Inquiries after the experiment showed that many participants discontinued their participation due to other priorities, such as work or social obligations. The experiment was completed by 10 men and 9 women with a mean age of 39 ($SD = 15.0$). There were 10 completed responses from VG participants and 9 from MG participants. The VG and MG groups did not differ significantly regarding sex, age and interest in gaming.

In the pre-experiment questionnaire (Q1), participants indicated their prior knowledge on a scale of 1 to 10. This resulted in a mean score of 4.2 ($SD = 2.4$) with no significant difference between groups.

4.2.3 Procedure

Participants had to register for the experiment by submitting an online consent form. After registration, each participant was automatically and randomly assigned to one of the Gameplay conditions.

The participants answered the online pre-experiment questionnaire before starting on the training. They were then asked to study the written materials and play the game, if applicable for their respective Gameplay condition. Voluntary players were free to decide if and how long they played, while mandatory players were told to spend a minimum of ten minutes playing. Participants could complete all parts of the training, at their convenience, through a web page with the experiment instructions and materials. They were free to study as long as they wished and proceed to the test when ready.

Figure 4.4 shows the experiment's procedure in relation to the materials.
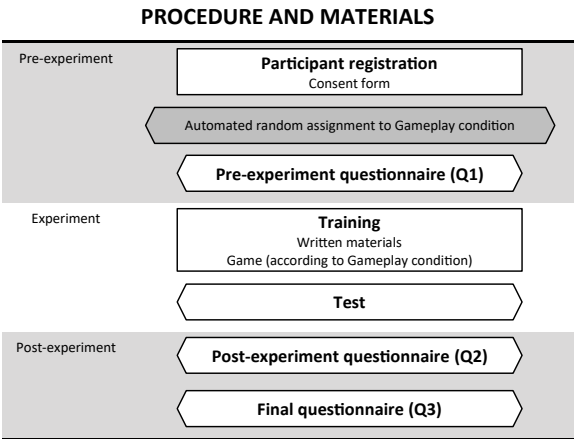
PROCEDURE AND MATERIALS

Figure 4.4: Schematic representation of the procedure and corresponding materials

4.2.4 Materials

The materials, developed for Experiment 1, were used in all three experiments. For Experiment 2 and Experiment 3 some adaptations were made. The complete set consisted of three parts.

- **A. Questionnaires.** A pre-experiment questionnaire (Q1), a post-experiment questionnaire (Q2), and a final questionnaire (Q3).
- **B. Training materials.** Written materials and a serious game.
- **C. Test.** A test with questions regarding cloud identification and risk assessment.

All materials (see Appendix B) were available online. The experiment's materials are shown in relation to the procedure in Figure 4.4. We will first discuss (A) the three questionnaires, followed by (B) the training materials, and (C) the test.

A. Questionnaires

The participants were presented with three questionnaires, which are briefly discussed below: a pre-experiment questionnaire (Q1) with a short assessment of prior knowledge at the time of registration, a post-experiment questionnaire (Q2) with questions about motivation directly after the experiment, and a final questionnaire (Q3) a few weeks after the experiment. All questionnaires are included in Appendix B.

Since no validated questions about voluntariness or enjoyment have been found in the literature, the questionnaires have been constructed specifically for the study.

In addition to multiple choice questions and 5- and 7-point Likert scale items [122], the questionnaire contains items with a 10-point scale. Such a scale is easily understood across age groups and education levels [86] and provides appropriate data for analysis

[35]. The use of an even scale avoids the neutral midpoint, forcing the participants to make a distinct choice for each item. Furthermore, the use of a 10-point scale is common in both customer satisfaction questionnaires and game reviews.

Pre-experiment questionnaire (Q1). After registration, participants were presented with the pre-experiment questionnaire with questions about (1) demographic information, (2) level of motivation, and (3) level of prior knowledge of aviation and meteorology.

Post-experiment questionnaire (Q2). After the test, participants were presented with the post-experiment questionnaire. This questionnaire solicited information about gaming preferences and personal motivation. The voluntary players were asked about the extent of the freedom of choice they experienced in choosing to play or not to play the game. The mandatory players were asked whether they would have played the game when given a choice. Upon completion of the test and the post-experiment questionnaire, participants were informed about the follow-up and about their chance of winning the gift certificate.

Final questionnaire (Q3). A few weeks after completion of the experiment, all participants were asked to answer a short, final questionnaire. The final questionnaire (Q3) for Experiment 1 (see p. 188 in Appendix B.3) contained 5 questions. The participants were asked about how they had heard about the experiment and their reason to participate. They were also asked to express their opinion about the experiment.

B. Training materials

Through a website, the participants were presented with two types of training materials: (1) written materials, and (2) a serious game. Depending on their assignment to a Gameplay condition, the participants received access to a specific version of the game. In both versions of the CloudAtlas game, the game was played in the exact same way (see the description of the game below). In the VG version of the game, voluntary players had access to the test and questionnaire (Q2) at any time. In contrast, the mandatory players had to play the MG version of the game for at least 10 minutes before they could continue to the test and questionnaire (Q2).

Written materials. The written materials consist of approximately 2000 words (see Appendix B.4). The written materials offer information about (1) cloud classification, (2) characteristics of the ten cloud types, (3) three possible hazards, and (4) the effect of clouds on aviation. The materials show drawings and photographs of different types of clouds. Both VG and MG players had unlimited access to the same set of text-based materials.

The CloudAtlas Game. The *CloudAtlas* game was designed to be played in an internet browser using the keyboard as the input device. Each individual game was relatively short. The game was designed to challenge the players to improve their high score, thus providing repeated exposure to the cloud types and their consequences.

Game environment. The game environment showed a side view of a simple landscape with a runway and the sky above it (Figure 4.5). On the left side was a small aircraft. From the right side, clouds and objects entered the screen. At the bottom of the screen, a dashboard provided information about the amount of fuel and oxygen available, the current game score, hazards and the weather conditions.

Gameplay. The goal of the game was to fly an aircraft as far as possible. The player had to adjust the altitude of the aircraft to avoid clouds and obstacles, or to land the aircraft if needed. The game ended when the player ran out of fuel or oxygen. The distance travelled translated into a game score.

Rules. The *CloudAtlas* game had eight main rules.

1. The aircraft had to take-off. It was not allowed to leave the aircraft on the runway.
2. Flying used up fuel. Flying at low altitude consumed more fuel than flying at high altitude.
3. Flying at high altitude consumed oxygen.
4. Flying through or under clouds exposed the aircraft to the hazards.
5. The chances of the hazards occurring and the intensity with which they occur, depend on the type of cloud, the weather conditions and a small random factor.
6. After an initial warning, increasingly more points were deducted from the score for unnecessary landings.
7. Collisions with objects immediately ended the game.
8. Running out of fuel or oxygen ended the game.

Resources. The player started each game with a limited supply of fuel and oxygen. During the game, the player could fly through boosters to receive extra fuel and oxygen.

Clouds. During flight, the player encountered ten types of clouds that were addressed in the written materials. Applying their knowledge about clouds and possible hazards, the players had to make decisions on how to respond. They could (1) fly through a cloud, (2) go over or under it, or (3) land the aircraft to wait for the danger to pass.

Hazards. Clouds could lead to three hazards to the aircraft: icing, turbulence and lightning. These hazards were visualised on the screen and had an effect on the game by increasing the aircraft's fuel consumption. Hazards could be avoided by flying at high altitude, but this required oxygen.

Objects. The player also encountered balloons and flocks of birds. Collisions had to be avoided because they ended the game.

Score. At the end of the game, the distance travelled with the aircraft translated into a game score. Picking up boosters during flight added to the score, while points were deducted for making unnecessary landings.

Figure 4.5:   CloudAtlas screenshot

C. Test

After playing the game, the participants could proceed to the test. The test consisted of eleven knowledge questions and seven application questions. The knowledge questions asked participants to reproduce cloud characteristics and recognise clouds from drawings and photographs.

In the application questions, participants applied their knowledge to a given situation. A picture of a game situation with an aircraft and a particular type of cloud was presented with four possible routes. Participants are asked to choose the best route, taking into consideration safety, comfort and efficiency, and to identify their reason or reasons for choosing the specific answer. An example of an application question is provided in Figure 4.6. In the example, Route A would be the best option. It would be unnecessary and inefficient to choose route D and land the aircraft. It would be unsafe to take route B or C, because of the risk of a collision with the birds. Route B may also cause some discomfort due to turbulence, and there is a risk of icing on the wings of the aircraft. Moreover, it is also necessary to look at the cloud further ahead. It would be better to pass over this cloud than to go under it, because of the risk of lightning.

All questions in the test had weights assigned to them. In general, application questions were considered to be more important. Therefore, they were assigned higher weights than knowledge questions. Test scores were calculated as the percentage of points earned out of the maximum.

The test makes up the first part of the post-experiment questionnaire (Q2), which can be found in Appendix B.2 on p. 176.
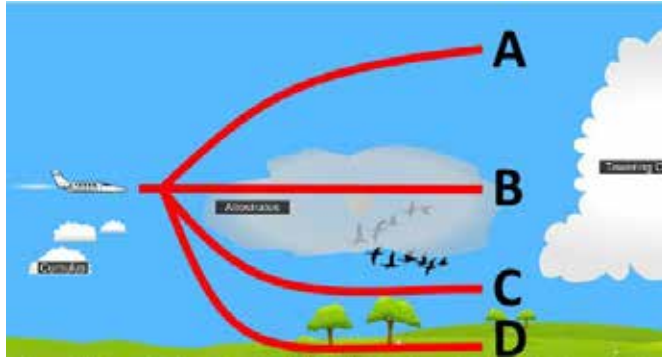
Figure 4.6:   Screenshot of CloudAtlas test item: application question

4.2.5   Results

In the first experiment, 19 participants (10 VG and 9 MG) completed the experiment by taking the test. The game was played by 16 of them, 3 VG participants chose not to play the game. Below, we discuss the results with regard to (A) the learning effect, and (B) the gameplay experience.

A. Learning effect

For the learning effect in Experiment 1, we will look at the results for Measurement 1: Game score, Measurement 2: Test score, and Measurement 3: Time spent playing the game.

Measurement 1: Game score. Game scores ranged from 721 to 4770. This has resulted in a large standard deviation for game score among the 10 voluntary players (VG) and the 9 mandatory players (MG). Table 4.2 shows the means and standard deviations on game scores. Contrary to our expectations, we did not find a significant difference between the voluntary and mandatory players on their performance in the game expressed in the game scores.

A t-test revealed that there were no significant differences in game score between male and female participants. However, gamers did achieve a higher game score than non-gamers ($F(1,17) = 8.4$, $p < 0.01$). Participants aged 40 and below scored significantly higher in the game ($F(1,17) = 15.6$, $p < 0.01$) than participants over the age of 40.

Measurement 2: Test score. Test scores ranged from 25 to 77. Table 4.2 shows the means and standard deviations on test scores. Contrary to our expectations, we did not find a significant difference between the voluntary and mandatory players on their performance in the test after playing the game.

A t-test revealed that there were neither significant differences in test score between male and female participants, nor was there a difference between gamers and non-gamers for test score. Participants aged 40 and below scored significantly higher on the test ($F(1,17) = 4.9$, $p < 0.05$) than participants over the age of 40.

Table 4.2: Means and SD of results for VG and MG conditions

Measure	Gameplay			
	VG n = 10		MG n = 9	
	M	SD	M	SD
Game score	1092	1085	2723	1332
Test score (%)	44.9	11.3	48.7	18.3

Note. In the VG group (n = 10), 3 participants chose not to play the game. Their game scores were 0.

Measurement 3: Time spent playing the game. We found a considerable variation in length of gameplay. In the VG group, 3 participants did not play at all, while 2 participants in the MG group played for more than half an hour. The number of tries varied from 0 to 22.

Figure 4.7 shows that MG participants ($M = 16.8$, $SD = 8.2$) played longer than VG participants ($M = 3.4$, $SD = 2.9$). This average difference of 13.4 minutes is significant ($F(1,17) = 23.5$, $p < 0.01$). The effect of the Gameplay condition on the amount of time played using prior motivation as a covariate was significant, $F(1,16) = 11.0$, $p < 0.01$. Contrary to what we expected based on the literature [26, 27] (see Subsection 2.5.3), mandatory players played longer, not shorter than voluntary players. A t-test revealed that there was neither a significant difference in time played between male and female players, nor a difference between gamers and non-gamers. However, females did have a lower average time per game attempt ($F(1,14) = 5.9$, $p < 0.05$). Participants over the age of 40 also had a lower average time per attempt than younger participants ($F(1,14) = 4.6$, $p < 0.05$).

We had anticipated seeing two subsets of players in both the VG and the MG conditions: (1) a subset with players who played only as long as required (less than 12 minutes), and (2) a subset with those who continued playing (more than 12 minutes). Table 4.3 shows counts and percentages for these subsets. In the MG group, 3 participants played less than 12 minutes, and the other 6 played longer. Surprisingly, all 10 players in the VG group played less than 12 minutes. Within the VG group, we also expected to find players who did not play at all, and players that only played to get an idea of the game by playing three tries or less. In total, 3 players did not play at all, and 4 players played three tries or less (Table 4.4).
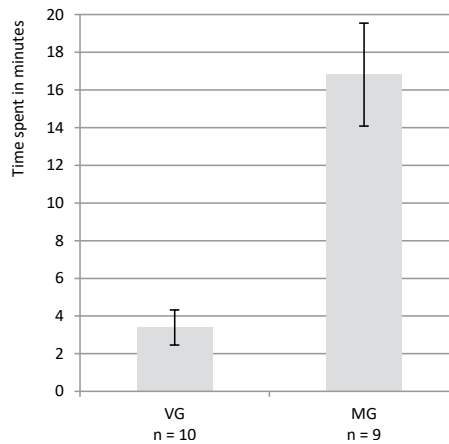
Figure 4.7: Means and SE for Time spent playing the game in Experiment 1

Table 4.3: Subsets in Voluntary and Mandatory Gameplay groups

Condition	Time spent playing	Gender		Gaming interest	
		Male n=10	Female n=9	Non-Gamer n=11	Gamer n=8
Voluntary (VG)	Less than 12 minutes	4	6	7	3
n=10	More than 12 minutes	0	0	0	0
Mandatory (MG)	Less than 12 minutes	1	2	2	1
n=9	More than 12 minutes	5	1	2	4

B. Gameplay experience

For the gameplay experience in Experiment 1, we will look at the results for Measurement 4: Enjoyment and Measurement 5: Obligation.

Measurement 4: Enjoyment. In the post-experiment questionnaire (Q2), all participants that played the game (n = 16) were asked how much they had enjoyed playing the game on a scale from 1 to 10 ($M = 6.6$, $SD = 1.6$). We found that younger participants enjoyed the game more than older participants ($F(1,17) = 9.0$, $p < 0.01$), and gamers enjoyed it more than non-gamers ($F(1,17) = 5.5$, $p < 0.05$).

Table 4.4: Subsets in Voluntary Gameplay group

Subset	Number of tries	Gender		Gaming interest	
		Male n=4	Female n=6	Non-Gamer n=7	Gamer n=3
Less than	No play	2	1	3	0
12 minutes	3 tries or less	2	2	2	2
n=10	4 tries or more	0	3	2	1

We had expected voluntary players to enjoy the game significantly more than mandatory players. Instead, we found that the difference was small and that the mandatory players (MG) even reported a slightly higher enjoyment (see Figure 4.8).
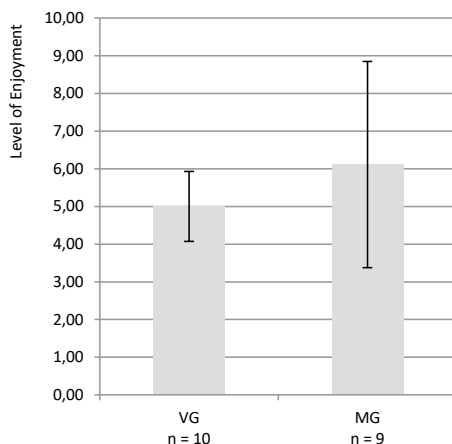

Figure 4.8:   Means and SE for Enjoyment in Experiment 1

Measurement 5:   Obligation.   Mandatory players were asked how they felt about being obliged to play the game for a minimum amount of time: bad, neutral or good. In general, they were neutral about this ($M = 2.1$, $SD = 0.8$). When asked whether they would play the game if they were given a choice, almost 78% of the mandatory players indicated they would. This percentage was higher than the percentage of voluntary players that actually chose to play the game, which was 70%. Furthermore, the duration of intended gameplay indicated by the mandatory players was higher than the average time played by the voluntary players, which was 3.4 minutes. Although these differences were not significant (as a result of the small sample of participants), they are opposite to expectation and as such a remarkable result.

A correlation for the data revealed that the feeling about being obliged to play and the decision to play the game if not mandatory were not significantly related, $r = 0.44$, n = 9, $p = 0.23$. A positive decision to play the game, if it were not mandatory, was not associated with a neutral or positive feeling about being obliged to play the game. Voluntary players were asked about the amount of freedom they experienced in choosing to play or not play the game on a scale from 1 to 10. The experienced levels of freedom ranged from 6 to 10, with a mean of 8.20 ($SD = 1.7$) and did not differ between gamers and non-gamers, male and female players or younger and older participants.

4.2.6 Discussion

The first experiment in the explorative study yielded interesting results with regard to the learning effect and the gameplay experience. In the results of Experiment 1, we also found results with regard to motivation and non-gamers. Below, we will discuss these four topics.(A) learning effect, (B) gameplay experience, (C) motivation, and (D) non-gamers.

A. Learning effect

This study aims to investigate the effect of voluntariness in a serious game on the learning effect. The learning effect of the serious game is measured by a test taken shortly after the training. We expected voluntary players to play the game longer and then perform better on the test than mandatory players. In reality, the data shows that mandatory players did spend more time playing the game. However, the time spent on training does not appear to be a factor. Performance does not differ statistically between the two groups. There are several candidate causes for this. We mention four of them. First, the group of voluntary players may have been able to extract knowledge from the game more efficiently than the mandatory players. Second, they may have been more successful in studying the written materials. Third, there may be design issues with the game or the test. Fourth, the game may not be as effective as expected, or the test may not be valid.

B. Gameplay experience

The second topic of interest is gameplay. Contrary to our expectations, voluntary players played for a shorter time than mandatory players and made fewer attempts. All voluntary players decided to quit playing the game within ten minutes. This raises the question of why they did so. Apparently, voluntary players did not become fully engaged in the game, even though they rate the game about the same for enjoyment as the mandatory players do. We observe that two-thirds of the mandatory players play more than two minutes beyond the ten-minute minimum. This shows that the game can be engaging. This outcome may indicate that a minimum time requirement is beneficial for gameplay, as it forces the participant not to give up at the first setback.

C. Motivation

In the pre-experiment question about their motivation to participate, players in the MG group indicated to be more motivated prior to the experiment than players in the VG group (one-way ANOVA: $F(1,17) = 9.3$, $p < 0.05$). At the time of answering this question, participants did not know yet which to group they had been assigned.

One-way analysis of variance controlled for motivation (ANCOVA) was used to control for the possible effects of the group difference that was found on motivation prior to the experiment. However, the effect of the covariate on Measurement 1 (game score), Measurement 2 (test score) and Measurement 3 (time spent playing) was not significant.

Participants may have been extrinsically motivated to participate in the experiment by the chance of winning a € 100 gift card. This extra motivation can be expected to have been equal between the voluntary (VG) and mandatory (MG) players. In line with the findings by Fulton and Schweitzer [67], we expected freedom of choice to motivate voluntary players and encourage them to accomplish better results.

Additionally, it would be understandable for a mandatory player to have a negative feeling about the obligation to play. However, voluntary players did neither do better on the test, nor did they score higher on the level of enjoyment than mandatory players. Mandatory players reported a neutral feeling about having to play the game for a minimum amount of time, not a negative one. The fact that one participates voluntarily in the experiment may change the way one feels about an obligation to play the game.

Alternatively, these outcomes may be caused by the small number of participants or the game design. Mandatory players even indicate that they would still play the game if it were not mandatory. Although the following results were not significant, considering the number of participants in the current study, they do indicate an interesting trend. The percentage of mandatory players, who said they would play the game without the obligation, was higher than the percentage of voluntary players who actually did. The gameplay duration estimated by the mandatory players was also higher than the time played by the voluntary players.

D. Non-gamers

We found that non-gamers played shorter and achieved lower scores than gamers. This may be indicative of the general gaming skills of this group. However, they did not perform worse on the test. These outcomes do not support the findings of Heeter et al. [84], who concluded that non-gamers are likely to be at a disadvantage in GBL. Also, the negative affect that Heeter et al. [84] found has not been established in the current study, even though non-gamers enjoyed the game less than gamers.

4.2.7   Section conclusion

Experiment 1 aimed to determine to what extent the learning effect and gameplay experience of a serious game are affected by the freedom to choose to play or not to play, i.e., whether playing the game is voluntary or mandatory. Due to the small number of participants who completed the experiment, no strong statistical conclusions can be drawn from the study.

We were surprised to find that voluntary players played shorter than mandatory players. As both groups reported equal enjoyment of the game, this difference does not need to be attributed to engagement. The outcomes suggest that mandatory players do not feel much pressure and that the obligation is mostly experienced as a stimulus.

We expected that using the game voluntarily as a learning tool would result in improved player performance in a test, in comparison to the results after mandatory gameplay (Measurement 1). We expected that voluntary play would have a positive effect on learning effect (RQ 2a). This result was not found.

With regard to gameplay (RQ 2b), we expected that voluntary players would enjoy the game more than mandatory players. Contrary to our expectations, we found that mandatory players played longer, not shorter (Measurement 3), and showed equal enjoyment (Measurement 4). Mandatory players were neutral about being obliged to play (Measurement 5).

This leads us to believe that mandatory gameplay in the *CloudAtlas* game does not ruin the enjoyment in the game. This contradicts the assumption of many game design theorists and practitioners that games need to be played voluntarily in order to be engaging, fun, and effective.

Recommendation

The findings of Experiment 1 indicate that the motivation of participants may be influenced by the way they have been recruited to participate, as well as by the incentives and rewards offered. We recommend that in the second and third experiment, the factor of motivation is taken into account by conducting the experiments in a more formal setting.

4.3 Experiment 2: Formal Setting

The second experiment served to determine whether voluntary gameplay in a learning tool in a formal setting would have the same effects, as we found in Experiment 1. In the formal setting, the primary motivation to participate would either be voluntary or mandatory. Experiment 2 had the same set-up as Experiment 1, but the experimental design was expanded with an NG. Moreover, the participants for the second experiment were recruited through institutions for higher education to be either voluntary (VP) or mandatory (MP) participants.

The new experimental design will be described first (Subsection 4.3.1), followed by the participants (Subsection 4.3.2), and the procedure (Subsection 4.3.3). Subsection 4.3.4 focuses on the adaptations that were made to the materials from the first experiment. Then, the results are reported in Subsection 4.3.5. A discussion is presented in Subsection 4.3.6, and conclusions on Experiment 2 are drawn in Subsection 4.3.7.

4.3.1 Design

The second experiment had a 2x3 experimental design (see Table 4.5). Participants were randomly assigned to one of the three Gameplay[2] groups.

1. **The voluntary gameplay group (VG)** in which players were free to choose how long to play the game or not to play the game at all.
2. **The mandatory gameplay group (MG)** in which players had to actively play the serious game for a minimum of 10 minutes.
3. **The No Game control group (NG)** in which players had no access to the game.

---

[2]Please note that in order to avoid confusion between participants and players, we use the abbreviation *P* for participation/participants and *G* for gameplay/players.

Remarks

1. The independent variables were (1) Participation and (2) Gameplay.
2. Participants were recruited either as voluntary (VP) or mandatory (MP) participants to the experiment. VP participants volunteered, while for MP participants the training was assigned as homework.
3. As an incentive, VP participants were offered a chance to win a € 100 gift certificate. They were informed that their chances of winning were related to completing all parts of the experiment, not to personal results.
4. However, VP participants receiving school credits for participating were not eligible for the gift certificate, neither were MP participants doing it as a homework assignment.
5. Our focus was on all seven measurements, viz. (1) game score, (2) test score, (3) time spent playing the game, (4) enjoyment, (5) obligation, (6) motivation, and (7) engagement.

Table 4.5: Participation and Gameplay conditions in Experiment 2

Participation	Gameplay		
	Voluntary (VG)	Mandatory (MG)	No Game (NG)
Voluntary (VP)	**Group VPVG** Volunteered to participate Free to choose to play	**Group VPMG** Volunteered to participate Minimum 10 minutes of gameplay	**Group VPNG** Volunteered to participate No access to game
Mandatory (MP)	**Group MPVG** Participation assigned Free to choose to play	**Group MPMG** Participation assigned Minimum 10 minutes of gameplay	**Group MPNG** Participation assigned No access to game

4.3.2   Participants

To create an obligation for students to participate, a formal learning setting was required in which lecturers assign the training as homework. For that reason, participants were recruited through (applied) universities. Students were informed that the experiment was related to aviation, but the focus on gaming remained undisclosed. Teachers and lecturers from sixteen faculties in twelve institutions were asked to assign participation in the experiment as a homework task to their students, creating a sense of obligation from the students toward their teacher. If a teacher was unable to assign homework for any reason, he informed the students about the experiment and invited them to participate without obligation.

Four teachers from four different institutions assigned homework to a total of approximately 90 students. One teacher posted the experiment on the ERAS network for Psychology students, who have to participate in experiments to get school credits. As

these students are free to choose the experiments in which to participate, they are considered to be voluntary participants in the current study. Thirteen teachers from seven different institutions invited over 1000 students to participate voluntarily.

A total of 93 participants completed the experiment. Homework had been assigned to 36 of them (MP), 31 participated for school credit (VP), and 26 volunteered (VP). Teachers did not report reasons for the high level of non-response. There were 42 men and 51 women with a mean age of 21.7 ($SD = 3.6$). The groups did not differ significantly in terms of sex, age, motivation prior to the experiment and number of gamers. However, of the 93 completed experiments, 19 participants spent less than 3 minutes studying the written materials combined with less than one minute playing the game, apart from the required time. They were considered non-legitimate participants, and their results were removed from the analyses. Another 7 participants were removed, as they were unable to play the game due to technical difficulties.

All in all, 67 participants completed the experiment; 33 men and 34 women with a mean age of 21.9 ($SD = 3.9$). In total, 45 students volunteered to participate, including the students who received school credit, and 22 students participated as part of a homework assignment. This will be referred to as voluntary (VP) and mandatory (MP) participation, respectively. Hence, Experiment 2 had 45 VP and 22 MP participants.

All participants were randomly assigned to a Gameplay group, resulting in a voluntary gameplay (VG) group of 22 participants, a mandatory gameplay (MG) group of 21, and a control group (NG) without access to the game of 24. The distribution of participants is shown in Table 4.6.

The test groups did not differ significantly in terms of sex, age, motivation prior to the experiment, and number of gamers.

Table 4.6: Distribution of participants for Participation and Gameplay in Experiment 2

Participation	Gameplay		
	VG n = 22	MG n = 21	NG n = 24
VP n = 45	**Group VPVG** 17	**Group VPMG** 15	**Group VPNG** 13
MP n = 22	**Group MPVG** 5	**Group MPMG** 6	**Group MPNG** 11

The participants' prior knowledge of aviation and meteorology was tested by a set of five questions in the pre-experiment questionnaire (Q1). After the test (Q2), participants were asked to indicate how much prior knowledge they had before the training. The average score on the prior knowledge assessment was 51.4 ($SD = 14.8$), and the average self-reported score on prior knowledge was 2.6 ($SD = 2.2$) on a scale of 1 to 10. These scores did not differ significantly between test groups. See Appendix B for the Q1 and Q2 questionnaires.

Furthermore, the two Participation groups did not differ in terms of sex, age, prior knowledge, and number of gamers. However, the VP participants (n = 45, $M = 7.5$, $SD = 1.1$) did report a significantly higher motivation prior to the experiment than the MP

Table 4.7: IMI subscales and items

IMI subscale		Items
1. Interest/Enjoyment		This task was fun to do
		I thought this was a boring task (R)
		This task did not hold my attention at all (R)
		I would describe this task as very interesting
		I thought this task was quite enjoyable
2. Perceived Competence		I am satisfied with my performance at this task
3. Effort/Importance		I put a lot of effort into this task
		I didn't try very hard to do well at this task (R)
		I tried very hard on this task
		It was important to me to do well at this task
		I didn't put much energy into this task (R)
4. Value/Usefulness		I believe this task could be of some value to me
		I think that doing this task is useful for training airline pilots
		I think this is an important task
	*	I believed playing the game could be beneficial to me
	*	I thought playing the game was an important activity
5. Pressure/Tension		I was very relaxed in doing this task (R)
		I was anxious while working on this task
		I felt pressured while doing this task
	*	I felt like I was expected to play the game (R)
	*	I believed I had a free choice about playing the game
6. Perceived Choice		I felt like it was not my own choice to do this task (R)
		I felt like I had to do this (R)
		I did this task because I wanted to

Items marked with * were added in the Q2 questionnaire for the VG group.

participants (n = 22, M = 5.5, SD = 2.5); $t(25.2)$ = -3.7, $p < 0.05$. One-way analysis of variance controlled for motivation (ANCOVA) was used to control for the possible effects of this difference.

4.3.3   Procedure

The second experiment's procedure was identical to that of the first experiment (see Figure 4.4 on p. 63). Participants could complete all parts online, at their convenience, through a web page with the experiment instructions and materials.

4.3.4   Adaptations to the materials

For Experiment 2, the materials from Experiment 1 were slightly adapted to accommodate the introduction of the control group (NG), as well as improve their usability. We will first discuss (A) the three questionnaires, followed by (B) the training materials, and (C) the test.

A. Questionnaires

A prior knowledge test was added to the pre-experiment questionnaire (Q1), and changes were made to the post-experiment (Q2) and final (Q3) questionnaires.

Pre-experiment questionnaire (Q1). Five questions about the cloud types and the associated risks were added to the pre-experiment questionnaire. These questions were used to establish an objective measurement of prior knowledge, in addition to the self-reported prior knowledge in the post-experiment questionnaire. Also, some questions were modified to reflect the changes in the experimental design (see p, 173, Appendix B).

Post-experiment questionnaire (Q2). Three changes were made to the post-experiment questionnaire. First, a post-experiment questionnaire was added for the no-gameplay (NG) control group. Second, the post-experiment questionnaire for all participants was expanded with a set of twenty questions (Table 4.7) from the Intrinsic Motivation Inventory (IMI) [91, 172]. The IMI is a multidimensional instrument to measure a participant's subjective experience with regard to the activity in an experiment. The current version of the IMI instrument offers six subscales: (1) Interest/Enjoyment, (2) Perceived Competence, (3) Effort/Importance, (4) Value/Usefulness, (5) Pressure/Tension, and (6) Perceived Choice. Intrinsic motivation is considered to be measured with the Interest/Enjoyment subscale. Perceived Competence and Perceived Choice are thought to be positive predictors of intrinsic motivation, while Pressure/Tension is a negative predictor. Relevant items from the IMI subscales can be selected and modified to fit specific activities. Third, an additional set of four questions relating to the IMI subscales of Value/Usefulness and Pressure/Tension was added to the post-experiment questionnaire for the VG group. In Table 4.7, these four questions are marked with *.

Final questionnaire (Q3). A new final questionnaire was constructed for Experiment 2. It contained 13 questions (see p. 189 in Appendix B). The participants were asked to indicate how they had been recruited, what their main reason to participate was, and how much they knew about the experiment beforehand. A set of statements about motivation were added, as well as a set of IMI statements [91, 172], in accordance with Q2.

B. Training materials

The website where the materials were made available was adapted to allow participants to be assigned to the NG control group. Furthermore, some changes were made to allow the use of Internet Explorer to play the game.

Written materials. No new information was added to the written materials, but they were slightly rearranged to improve the presentation. Also, learning objectives were added, and the navigation through the pages was made more evident, to give participants more control over their learning process. The written materials are presented in Appendix B.4.

The CloudAtlas Game. No changes were made to the game itself.

C. Test

The test was unchanged, except for a few minor textual corrections.

4.3.5   Results

In total, 67 participants completed the experiment by using the presented training materials, taking the test and answering the pre- and post-experiment questionnaires. The game was part of the training materials of 43 participants in the experiment (22 VG, 21 MG), of whom 37 played the game. A total of 6 participants assigned to the VG group chose not to play the game. Below, we discuss the results with regard to (A) the learning effect, and (B) the gameplay experience.

A. Learning effect

For the learning effect in Experiment 2, we will look at the results for Measurement 1: Game score, Measurement 2: Test score, and Measurement 3: Time spent playing the game.

Measurement 1: Game score. Game scores ranged from 254 to 4060. Table 4.8 shows the means and standard deviations on game score.

Measurement 2: Test score. Overall test scores ranged from 24 to 82. For a more detailed analysis, the overall test results of all participants were split into a score for knowledge questions and one for application questions. Means and standard errors of test scores are visualised in Figure 4.9. No significant effects of Participation or Gameplay on the test scores were found. Table 4.8 shows the means and standard deviations on test score.
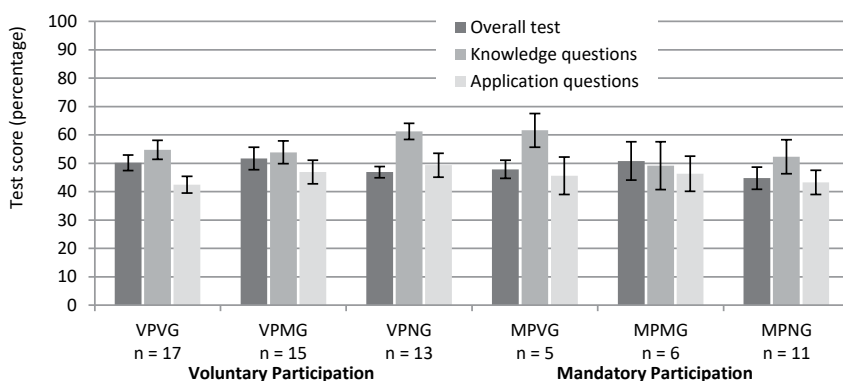

Figure 4.9:   Means and SE for Test score in Experiment 2

Measurement 3: Time spent playing the game. The time spent playing varied widely as 6 participants in the VG group did not play, while in the MG group, 5 participants played for more than 15 minutes. Figure 4.10 shows that MG participants spent more time playing the game than VG participants. On average, they spent 11.4

Table 4.8: Means and SD of results for Participation and Gameplay conditions in Experiment 2

| Participation		Gameplay								
		VG n = 22		MG n = 21		NG n = 24		Total n = 67		
		M	SD	M	SD	M	SD	M	SD	
VP n = 45		**Group VPVG** n = 17		**Group VPMG** n = 15		**Group VPNG** n = 13		**Total** n = 45		
	Game score	698	605	2343	735	N/A	N/A	1047	1120	
	Test score	50.2	11.4	51.7	15.3	46.9	7.1	49.2	12.4	
	Test - Knowledge	54.8	13.9	53.9	15.4	61.2	10.2	56.4	13.3	
	Test - Application	42.4	12.1	46.9	16.1	49.3	15.1	46.4	14.9	
MP n = 22		**Group MPVG** n = 5		**Group MPMG** n = 6		**Group MPNG** n = 11		**Total** n = 22		
	Game score	537	826	2705	949	N/A	N/A	855	1308	
	Test score	47.9	7.1	50.8	16.5	44.8	13.1	48.3	11.7	
	Test - Knowledge	61.6	13.4	49.2	20.6	52.3	19.8	53.3	18.9	
	Test - Application	45.6	14.7	46.3	15.2	43.3	14.1	43.6	12.5	
Total n = 67		**VG** n = 22		**MG** n = 21		**NG** n = 24				
	Game score	662	643	2446	795	N/A	N/A			
	Test score	49.6	10.5	51.4	15.2	45.9	10.1			
	Test - Knowledge	56.3	13.8	52.5	16.7	57.1	15.7			
	Test - Application	43.1	12.5	46.8	15.5	46.5	14.6			

Note. In the VG group (n = 22), 6 participants chose not to play the game. Their game scores were 0.

minutes more. The main part of this difference can be explained by the required minimum of 10 minutes of gameplay for the MG group. If we consider the time that is played beyond the 10 minutes to be voluntary, on average, MG participants show longer voluntary play than VG participants, but this difference is not significant ($p = 0.17$).

B. Gameplay experience

For the gameplay experience in Experiment 2, we will look at the results for Measurement 4: Enjoyment, Measurement 5: Obligation, and Measurement 6: Motivation.

Measurement 4: Enjoyment. Besides the quantitative data of the game scores and test scores, we have focused on the possible acceptance of a serious game in a mandatory setting. For this purpose, we have measured enjoyment of the game, opinion about being obliged to play, and willingness of mandatory players to play voluntarily. A total of 35 participants played the game (17 VPVG, 15 VPMG, 5 MPVG, 6 MPMG). Overall, they rated their enjoyment on a scale of 1 to 10 (n = 35, $M = 6.3$, $SD = 2.2$). As in the first experiment, we found that MG participants enjoy the game as much as VG participants (Figure 4.11). The same is true for MP participants. They indicate as much enjoyment in the game as the VP participants. In fact, both the MG and the MP groups give a higher score for enjoyment, although this is not a significant difference.
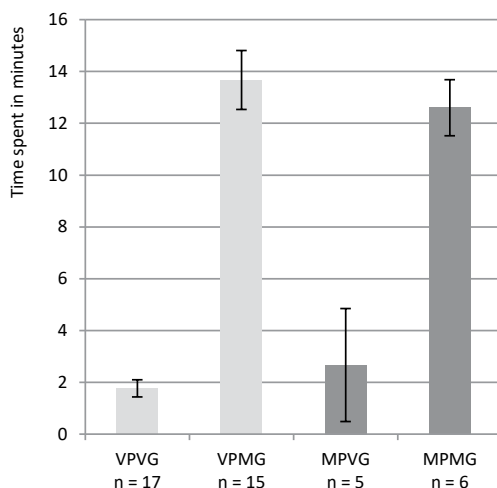
Figure 4.10: Means and SE for Time spent playing the game in Experiment 2
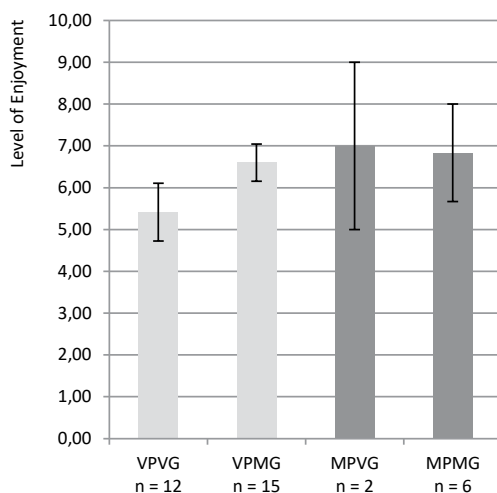
Figure 4.11: Means and SE for Enjoyment in Experiment 2

Measurement 5: Obligation. MG participants were asked how they felt about having
to play the game for at least 10 minutes: bad, neutral or good. Overall, they were a
bit negative in the second experiment (n = 21, $M = 1.7$, $SD = 0.6$). No difference
was found between the voluntary (VP) and mandatory (MP) participants. In the
post-experiment questionnaire (Q2), MG participants were asked if they would play
the game voluntarily. The NG participants were asked a similar question after they
were informed that other participants had played a game. These questions offered
four choices: (1) no, (2) probably not, (3) probably yes, and (4) yes. Table 4.9

shows the percentages of VG participants who actually played the game combined with the percentage of MG and NG participants expressing they would play the game voluntarily. An independent-samples t-test shows that the willingness of MG and NG participants does not differ significantly from the actual percentage of gameplay of the VG participants. The data reveals that less VG participants who were obliged to participate (MPVG, 40%) played the game than VG participants who volunteered to participate (VPVG, 82%). However, this difference is not significant ($p = 0.07$).

Measurement 6: Motivation. In the pre-experiment questionnaire (Q1), participants were asked to indicate their motivation to participate in the experiment on a scale of 1 to 10 (Figure 4.12). An independent-samples t-test revealed a significant difference. The motivation reported by VP participants is significantly higher than that of MP participants, $t(25.6) = 3.4$, $p < 0.01$. This is in accordance with the general expectation that volunteers are more motivated.

The post-experiment questionnaire (Q2) contained a set of twenty questions in regard to motivation, based on the IMI scales [91, 172], as described in Subsection 4.3.4. Two-way ANOVA tests were conducted to examine the influence of Participation and Test Group on the six IMI subscales. The significance levels are shown in Table 4.10. As was to be expected, a simple main effect showed that VP participants reported experiencing significantly more choice than MP participants ($p < 0.01$). No interaction effects or other main simple effects have been found.

Table 4.9: Actual gameplay versus reported willingness to play the game voluntarily

Participation	Actual gameplay		Reported willingness to play			
	VG n = 22		MG n = 21		NG n = 24	
VP n = 45	n = 17	82%	n = 15	67%	n = 13	77%
MP n = 22	n = 5	40%	n = 6	33%	n = 11	82%
Total	n = 22	73%	n = 21	57%	n = 24	79%

Table 4.10: Significance levels of between-subject effects for the IMI subscales

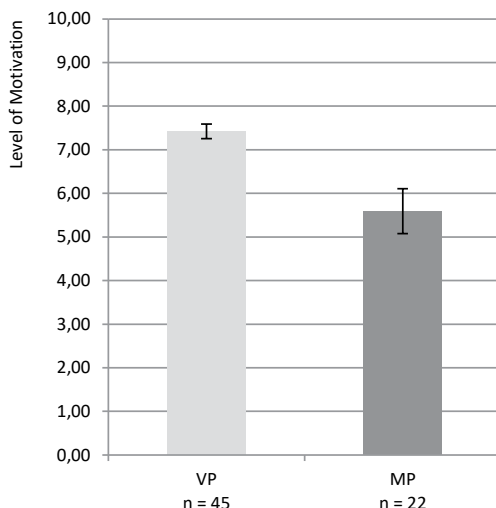| IMI subscale	Between Subjects Effects		
	Participation (VP, MP)	Gameplay (VG, MG, NG)	Interaction of Participation and Gameplay
Interest	.45	.85	.57
Pressure	.06	.49	.12
Choice	.00*	.48	.43
Value	.47	.98	.25
Effort	.80	.88	.39
Competence	.24	.80	.58

Note. * $p < 0.01$

Figure 4.12:   Means and SE for Motivation in Experiment 2

4.3.6   Discussion

The outcomes of Experiment 2 confirm what we have found in Experiment 1. Again, the mandatory (MP) participants and players did not seem to experience a negative effect from the obligation. No new surprises were encountered, but there were findings with regard to the incentives and rewards, and cultural aspects. Below, we will discuss (A) learning effect, (B) gameplay experience, (C) motivation, (D) non-gamers, (E) incentives and rewards, and (F) cultural aspects.

A. Learning effect

The study sought to investigate the effect of voluntariness on the learning effect of a serious game. A test, taken shortly after the training, measured the learning effect of the serious game, in combination with the game scores and the time spent playing the game. The data showed that mandatory players (MG) spent more time playing the game. However, performance on the test does not differ statistically between the test groups, similar to what we found in Experiment 1. Therefore, time spent on training does not appear to be a factor. The topic of the training and the game may have been outside the area of interest for most participants, or the game may not be as effective as expected.

B. Gameplay experience

Almost half of the voluntary players (VG) chose not to play the game at all, and the participants who did play played for a shorter time than mandatory players (MG). This difference can be explained by the required minimum of 10 minutes of gameplay for the MG group. Only one VG participant played for more than 10 minutes, and all others played

less than 4 minutes. Even after correcting for the required minimum of 10 minutes, MG participants played significantly longer. This may indicate that MG participants become more engaged in the game than VG participants. This seems to be unrelated to the fun of the game, as participants in the VG and MG groups report an equal score on enjoyment.

On the one hand, mandatory (MG) players from the mandatory participation (MP) group played, almost two minutes longer than MG participants from the VP group; an interesting although not significant difference. No explanation has been found for this difference, but it could be the effect of the obligation toward the lecturer. On the other hand, the VG participants in the MP group play half a minute shorter than the VG participants in the VP group, on a total playtime of two minutes. It seems that, contrary to the MG participants, VG participants do not feel obliged to keep playing.

C. Motivation

Motivation plays an essential role in this study, both intrinsic and extrinsic motivation. We found that VP participants were more motivated than MP participants (Measurement 7). A possible explanation for this is that participants with low motivation may not participate if they have the choice (as is the case in VP), but in MP, they are obliged to participate, which in effect may reduce the overall score for motivation.

We expected freedom of choice to motivate voluntary players and encourage them to accomplish better results. However, voluntary players (VG) did neither do better in the game and on the test nor did they score higher on the level of enjoyment than mandatory players (MG), irrespective of whether they participated voluntarily (VP) or as part of a homework assignment (MP).

Additionally, it would be understandable for a player to experience negative feelings when obliged to play the game, but mandatory players (MG) reported a neutral feeling about having to play the game for a minimum amount of time. The data suggest that a negative feeling about the minimum time is neither related to voluntary (VP) or mandatory participation (MP) nor is it related to a participant's motivation to participate in the experiment.

The majority of both mandatory players (MG) and control group participants (NG), indicated that they would play the game if they were offered a choice. The percentage of no-gameplay (NG) participants, who said they would play the game without the obligation, was higher than the percentage of voluntary players (VG) who actually did play the game.

D. Non-gamers

In Experiment 2, we did not examine the differences between gamers and non-gamers, as it was not the focus of our study. Besides, the results of Experiment 1 had shown that the difference did not have an effect on our measurements.

E. Incentives and rewards

In the experiment, incentives and rewards were offered to students for their participation. Similar to the first experiment, voluntary participants (VP) who were invited to participate by their teacher, were offered a chance of winning a € 100 gift card, while participants

through the ERAS network received school credit and were not eligible to win the gift card. Based on the individual lecturer's choices, some mandatory participants (MP) were offered the chance of winning the gift card, some received school credit and others did not get any reward.

However, offering a reward may have had an adverse effect on participants. A meta-analysis of studies on the effect of extrinsic rewards on intrinsic motivation has shown that tangible rewards for interesting tasks may diminish the intrinsic motivation [47, 48].

F. Cultural aspects

The participants in the voluntary participation (VP) group were all enrolled in schools in the Netherlands, whereas the students in the mandatory participation (MP) group come from schools in the Netherlands, Thailand, and Lebanon. As these countries have different cultures, cultural differences may have had an effect on the outcomes of the experiment. Among other things, students in Thailand and Lebanon may have a more formal relationship with their teachers than Dutch students, possibly resulting in a stronger sense of obligation.

4.3.7   Section conclusion

Based on the test scores (Measurement 1) on the pre-test and post-test, we conclude that the *CloudAtlas* game does not have a significant learning effect. Therefore, it was not possible to determine the effect of voluntary and mandatory play on the learning effect. Hence, Experiment 2 does not provide an answer to RQ 2a.

Experiment 2 confirms the outcomes of Experiment 1 with regard to RQ 2b: the obligation to play the game, and even the obligation to participate in the experiment, appears to have no negative effect on the time spent playing (Measurement 3) and the enjoyment of the game (Measurement 4). We found that a little coercion increases the time spent in the game, which in turn may improve the learning effect. Moreover, the results indicate that mandatory gameplay is just as much fun as voluntary gameplay.

Recommendation

Despite our efforts, the distinction between voluntary (VP) and mandatory participation (MP) was insufficient, partly because of the rewards offered. We recommend that in the third experiment the voluntary participants (VP) should volunteer out of their personal interest without any rewards, and the mandatory participants (MP) should be instructed to participate by a person who has a strong influence on their education or career.

4.4   Experiment 3: Formal setting with mandatory participation

The third experiment was set up to determine, once again, the effect of voluntarily using a game as a learning tool on the learning outcome and gameplay experience of the serious game.

There were four differences with the previous experiments: (1) a stronger distinction between voluntary and mandatory participation, (2) participants for whom the topic of the training materials was more relevant, (3) the NG control group was eliminated, and (4) no incentives or rewards were offered. Participants were divided into four test groups based on Participation and Gameplay.

The new experimental design is discussed first (Subsection 4.4.1), followed by the participants (Subsection 4.4.2) and the procedure (Subsection 4.4.3). Subsection 4.4.4 discusses the adaptations that were made to the materials from the previous experiments. Then, the results are presented in Subsection 4.4.5. The discussion is in Subsection 4.4.6, and the conclusion in Subsection 4.4.7.

4.4.1 Design

The experimental design for the third experiment was similar to that of the second experiment, but with four conditions as a result of eliminating the NG control group (Table 4.11). From Experiment 1 and 2, we concluded that the serious game did not have a learning effect. Therefore, there was no need to have an NG control group in Experiment 3 to compare the effect of the game with that of the written materials.

The independent variables were (1) Participation and (2) Gameplay. Participants were recruited either as voluntary (VP) or mandatory (MP) participants to the experiment. In both Participation groups, participants were assigned to one of two Gameplay[3] groups.

1. **The voluntary gameplay group (VG)** in which players were free to choose how long to play the game or not to play the game at all.
2. **The mandatory gameplay group (MG)** in which players had to actively play the serious game for a minimum of 10 minutes.

Our focus was on all seven measurements, viz. (1) game score, (2) test score, (3) time spent playing the game, (4) enjoyment, (5) engagement, (6) obligation, and (7) motivation.

Table 4.11: Participation and Gameplay conditions in Experiment 3

Participation	Gameplay	
	Voluntary (VG)	Mandatory (MG)
Voluntary (VP)	**Group VPVG** Volunteered to participate Free to choose to play	**Group VPMG** Volunteered to participate Minimum 10 minutes of gameplay
Mandatory (MP)	**Group MPVG** Participation assigned Free to choose to play	**Group MPMG** Participation assigned Minimum 10 minutes of gameplay

---

[3]Please note that in order to avoid confusion between participants and players, we use the abbreviation *P* for participation/participants and *G* for gameplay/players.

4.4.2   Participants

For the third experiment, an educational population was chosen for whom the topic of the training was more relevant, in order to have a higher level of *perceived utility* [34, 121, 168, 170] (see also Definition 4.8 on p. 60). A sense of utility is believed to have a positive effect on a person's motivation [34, 203]. Recruitment was aimed at persons working on obtaining their PPL and licensed pilots with a small amount of flight experience. In contrast to the first and second experiment, the third experiment offered no incentives or rewards to any of the participants.

To create the required formal learning setting, we contacted the Chief Flight Instructors of four Dutch flight academies. The students in their academies wish to become licensed pilots with the connected airlines. Therefore, students will feel obliged to participate in activities assigned by the flight instructors. A total of 101 students and 11 pilots were instructed to participate; 70 participants registered for the experiment and 54 of them completed the test.

Voluntary participants (VP) were recruited through (1) the Royal Netherlands Aeronautical Association (KNVvL), (2) the Airwork bulletin board, and (3) flight schools for leisure aviation. An announcement was included in the KNVvL newsletter in May and July 2017, and the announcement was also published on their website. The same announcement was used on the Airwork bulletin board in May and July 2017. In May 2017, some PPL pilots and flight instructors forwarded the invitation to participate in the experiment to their flight school members. A total of 81 VP registered, 34 of whom completed the experiment.

Initially, the experiment was completed by 89 participants; 34 were voluntary participants (VP), 55 were recruited through the flight academies (mandatory participants, MP). There were 85 men and 4 women with a mean age of 33.3 ($SD = 16.3$).

However, the results of 12 participants were excluded from the analyses. In the VG condition, 9 participants reported that they were unable to play the game due to technical difficulties. There were 3 participants (2 MP, 1 VP) who finished the test without spending any time on the game and the written materials. They were considered non-legitimate participants. Another 2 participants spent less than 3 minutes on the game and the written materials combined, but they were not removed, as their decision to skip part of the materials may have been based on their personal assessment of their prior knowledge of the topic.

Hence, 77 participants completed the experiment; 73 men and 4 women with a mean age of 32.0 ($SD = 16.0$). The distribution of participants is shown in Table 4.12.

Although recruitment was focused on pilots in training and pilots with little experience, 4 pilots participated with a larger number of flight hours and a background as a professional pilot. However, there was no indication that these 4 pilots deviated strongly from the other participants in other aspects.

The test groups differed in a number of ways. The voluntary (VP) participants (n = 25, $M = 50.4$, $SD = 14.4$) were older than those in the MP group (n = 52, $M = 23.1$, $SD = 6.0$); $t(28.1) = -9.1$, $p < 0.01$. Within the VP group, post hoc testing using the Bonferroni correction revealed that Group VPVG (n = 12, $M = 61.6$, $SD = 8.5$) was

Table 4.12: Distribution of participants for Participation and Gameplay in Experiment 3

Participation	Gameplay	
	Voluntary (VG) n = 38	Mandatory (MG) n = 39
Voluntary (VP) n = 25	**Group VPVG** 12	**Group VGMG** 13
Mandatory (MP) n = 52	**Group MPVG** 26	**Group MPMG** 26

older than Group VPMG (n = 13, M = 40.1, SD = 10.5); $p < 0.01$. However, ANCOVA testing showed that age as a covariate had no significant effect on the outcomes. Hence, the covariate was removed from the analyses.

The VP participants (n = 25, M = 7.8, SD = 1.1) reported a significantly higher motivation prior to the experiment than the MP participants (n = 52, M = 7.2, SD = 1.1); $t(75) = -2.4$, $p < 0.05$. Post hoc testing using the LSD correction showed that only Group VPMG was more motivated than Group MPMG ($p < 0.01$) and Group MPVG ($p < 0.05$).

The participants' prior knowledge on meteorology was tested with a set of five questions in the pre-experiment questionnaire (Q1). After the test (Q2), participants were asked to indicate on a 10-point scale how much prior knowledge they had before the training. The average score on the prior knowledge assessment was 67.5 (n = 77, SD = 12.1), and the average self-reported score on prior knowledge was 6.1 (n = 77, SD = 1.8) on a scale of 1 to 10. No significant differences were found between the VP and MP groups, the VG and MG groups, or the four separate test groups.

4.4.3 Procedure

The third experiment's procedure was identical to that of the first and second experiments (see Figure 4.4), except for the automatic assignment to the Gameplay condition. In the third experiment, participants were automatically assigned to a Gameplay condition according to a schedule (Table 4.13) instead of randomised assignment. This method was chosen to control the number of completed participations in each group. However, the assignment was blinded and was not influenced by the individual participants.

Table 4.13: Assignment of registered participants to Gameplay condition in Experiment 3

Registration	Assignment
1-10	MG
11-20	VG
21-30	MG
31-40	VG
41 and following (uneven)	MG
42 and following (even)	VG

4.4.4   Adaptations to the materials

For Experiment 3, the materials from Experiment 2 were adapted to accommodate the recruitment of participants, and for the training materials to have a closer fit with the study materials used in flight schools. We added a third questionnaire, to be answered by the participants after completion of the experiment. We will first discuss (A) the three questionnaires, followed by (B) the training materials, and (C) the test.

A. Questionnaires

Pre-experiment questionnaire (Q1). The third experiment used the pre-experiment questionnaire from the second experiment, with the exception of six questions that had been added before to identify participants as voluntary (VP) or mandatory participants (MP). In the third experiment, this distinction was made through the experiment website.

Post-experiment questionnaire (Q2). Questions that were specific to the participants of the second experiment were replaced with questions about pilot licenses, flight training and meteorology exams. All other questions were identical to the previous versions of the post-experiment questionnaire.

Final questionnaire (Q3). The final questionnaire was adapted for Experiment 3 (see p. 191 in Appendix B). It contained 7 questions. Questions about the recruitment that were no longer relevant were removed. The participants were asked about the amount of obligation that had experienced and how they had used the learning materials. They were asked to express their opinion about the experiment. Furthermore, the questionnaire contained a set of IMI statements [91, 172].

B. Training materials

Website. Changes were made to the website to provide separate URLs for voluntary (VP) and mandatory participants (MP). The flight instructors received personal URLs to hand out to their students, linking their students to them. A generic URL was used to invite voluntary participants (VP).

Written materials. The written materials were adapted to provide a closer match with the Meteorological Theory study materials [13] for PPL courses. Also, some small typographical errors were corrected. The written materials are presented in Appendix B.4.

CloudAtlas game. For Experiment 3, the game was adapted to give the player feedback on dangerous actions. A player will be warned for the possible consequences when flying through a dangerous cloud for the first and second time (see Figure 4.13). The third and following incidents will not trigger the feedback anymore. Furthermore, the adapted game shows the names of the clouds together with the abbreviations that are commonly used in aviation.

Figure 4.13: CloudAtlas screenshot: feedback on dangerous action

C. Test

In the third experiment, the same test was used as in the second experiment. No changes were made.

4.4.5 Results

After removing the results of the participant's who completed the experiment in an invalid way, 77 completed participations remained. Of the 38 VG participants, 3 did not play the game. Below, we discuss the results with regard to (A) the learning effect, and (B) the gameplay experience.

A. Learning effect

For the learning effect in Experiment 3, we will look at the results for Measurement 1: Game score, Measurement 2: Test score, and Measurement 3: Time spent playing the game.

Measurement 1: Game score. A total of 74 participants played the game, and their game scores ranged from 462 to 4978. A significant main effect of both Participation ($F(1,70) = 4.5$, $p < 0.05$) and Gameplay ($F(1,70) = 29.2$, $p < 0.01$) was found. However, the interaction effect of Participation and Gameplay on game score was not significant. Table 4.14 shows the means and standard deviations on game scores.

Table 4.14: Means and SD of results for Participation and Gameplay conditions in Experiment 3

Participation		Gameplay					
		VG n = 38		MG n = 39		Total n = 77	
		M	SD	M	SD	M	SD
		Group VPVG		**Group VPMG**		**Total**	
		n = 12		n = 13		n = 25	
	Game score	899	675	2564	996	1765	1194
VP	Test score	58.0	14.4	58.5	11.9	58.3	12.9
n = 25	Test - Knowledge	62.8	9.8	62.8	15.4	62.8	12.7
	Test - Application	55.0	19.5	55.7	15.1	55.4	17.0
		Group MPVG		**Group MPMG**		**Total**	
		n = 26		n = 26		n = 52	
	Game score	1764	947	2749	891	2256	1038
MP	Test score	54.3	8.2	56.0	11.2	55.2	9.7
n = 52	Test - Knowledge	62.9	9.3	62.1	11.7	62.5	10.5
	Test - Application	48.8	11.7	52.2	15.5	50.5	13.7
		VG		**MG**			
		n = 38		n = 39			
	Game score	1491	953	2687	918		
Total	Test score	55.5	10.5	56.8	11.3		
n = 77	Test - Knowledge	62.9	9.3	62.4	12.8		
	Test - Application	50.8	14.6	53.3	15.3		

Note. In the VG group (n = 38), 3 participants chose not to play the game. Their game scores were 0.

Measurement 2: Test score. Overall, test scores ranged from 28 to 83. For a more detailed analysis, the overall test results of all participants were split into a score for knowledge questions and one for application questions (Figure 4.14). No significant effects were found for the overall test score, the score for knowledge questions, and the score for application questions.

Table 4.14 shows the means and standard deviations on test scores. As in the previous experiments, no significant effects of Participation or Gameplay on test scores were found.

Measurement 3: Time spent playing the game. Gameplay length varied widely (Figure 4.15). In the VG group, 3 participants did not play the game at all, and 19 participants played the game three times or less. In the MG group, 10 participants played for more than 15 minutes. None of the VG participants played more than 13 minutes. In fact, only 10 VG participants followed the recommendation on the website and played for a minimum of ten minutes. All but one of them were part of the MP group. On average, MG participants spent 9 minutes more playing the game than VG participants. This difference can be explained by the required minimum of 10 minutes of gameplay for the MG group. Two-way analysis of variance shows a significant effect of voluntary and mandatory gameplay on the length of gameplay, $F(1,73) = 113.8$, $p < 0.01$. Furthermore, there is a significant effect of the interaction of Participation and Gameplay on the length of gameplay,

F(1,73) = 11.1, $p < 0.01$. Looking at voluntary playtime only, by subtracting the mandatory ten minutes from the total time played for MG participants, a significant effect of the interaction of Participation and Gameplay on the length of gameplay remains, F(1,73) = 11.1, $p < 0.01$.

The same effect is shown when comparing the voluntary playtime for the four test groups. Post hoc testing using the Bonferroni correction for the four test groups shows significant differences between Group MPMG and Group MPVG ($p < 0.01$) and Group MPVG and Group VPVG ($p < 0.05$).
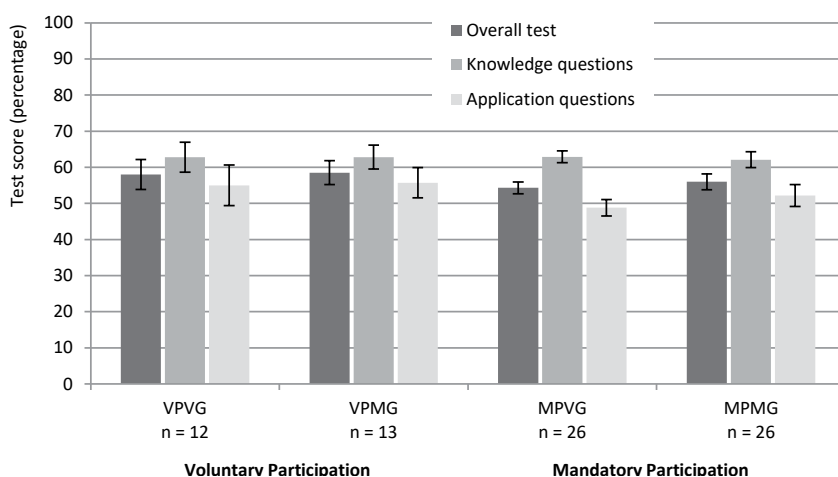

Figure 4.14: Means and SE for Test score in VP and MP conditions in Experiment 3

B. Gameplay experience

For the gameplay experience in Experiment 3, we will look at the results for Measurement 4: Enjoyment, Measurement 5: Obligation, Measurement 6: Motivation, and Measurement 7: Engagement.

Measurement 4: Enjoyment. The 74 participants that played the game rated their enjoyment on a scale of 1 to 10 ($M = 5.8$, $SD = 2.0$). Figure 4.16 shows the means and standard error for the enjoyment scores. One-way ANOVA shows no significant effects between the four test groups, nor between the VP and MP groups, and the VG and MG groups.

As in the first and second experiment, MG participants rate their enjoyment equally as VG participants. The same is true for MP participants. They indicate as much enjoyment in the game as the VP participants. In fact, the participants in the MP groups give a higher score for enjoyment than the VP participants, although this is not a significant difference.
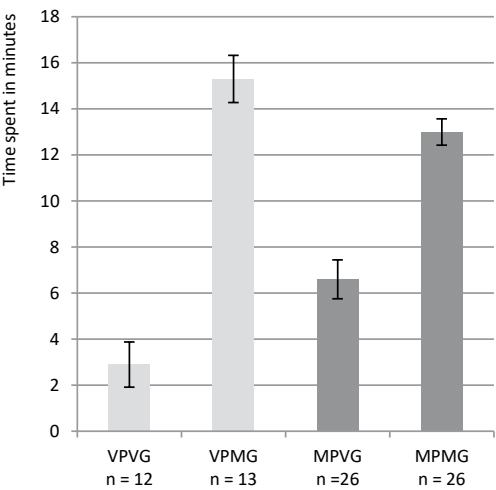
Figure 4.15: Means and SE for Time spent playing the game in Experiment 3
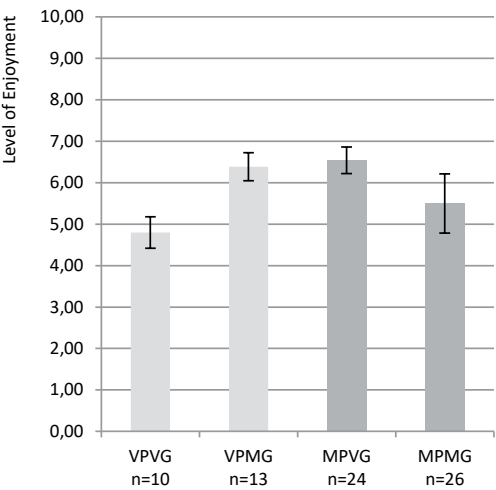
Figure 4.16: Means and SE for Enjoyment in Experiment 3

Measurement 5: Obligation. MG participants were asked how they felt about having to play the game for at least 10 minutes: bad, neutral or good. Overall, MG participants were a bit negative in the experiment (n = 39, M = 1.8, SD = 0.7). An independent-samples t-test shows that the participants in Group VPMG are significantly more positive than the participants in Group MPMG; $t(37)$ = -3.8, $p < 0.05$.

Measurement 6: Motivation. In the pre-experiment questionnaire (Q1), participants were asked to indicate their motivation to participate in the experiment on a scale of 1 to 10. Table 4.15 shows the means and standard deviations. A t-test revealed that the motivation reported by VP participants is significantly higher than that of MP participants, $t(75)$ = -2.4, $p < 0.05$.

The post-experiment questionnaire (Q2) contained a set of twenty questions in regard to motivation, based on the IMI subscales [91, 172]. Two-way ANOVA tests were conducted to examine the influence of Participation and Gameplay on the six IMI subscales. The significance levels are shown in Table 4.16.

Simple main effects showed that VP participants reported experiencing significantly more interest than MP participants, and VG participants reported significantly more choice than MG participants ($p < 0.05$). No interaction effects or other main simple effects have been found.

Table 4.15: Levels of Motivation separated for the Participation conditions in Experiment 3

Participation		Gameplay					
		VG n = 38		MG n = 39		Total n = 77	
		M	*SD*	*M*	*SD*	*M*	*SD*
		Group VPVG		**Group VPMG**		**VP**	
		n = 12		n = 13		n = 25	
VP n=25	Motivation	7.5	0.7	8.1	1.3	7.8	1.1
		Group MPVG		**Group MPMG**		**MP**	
		n = 26		n = 26		n = 52	
MP n=52	Motivation	7.2	0.9	7.1	1.3	7.2	1.1
		VG		**MG**		**Total**	
		n = 38		n = 39		n = 77	
Total n=77	Motivation	7.3	0.8	7.4	1.4	7.4	1.1

Measurement 7: Engagement. All participants rated their engagement within the entire training on a scale of 1 to 10 (n = 77, M = 7.0, SD = 1.6). Table 4.17 shows the means and standard deviations for the engagement scores. One-way ANOVA shows no significant effects between the four test groups. However, a one-way ANOVA showed a significant difference between the VP and MP groups; $F(1,75)$ = 4.9, $p < 0.05$. The VP participants reported a higher engagement.

Table 4.16: Significance levels of between-subject effects for the IMI subscales in Experiment 3

IMI subscale	Between Subject Effects		
	Participation (VP, MP)	Gameplay (VG, MG)	Interaction of Participation and Gameplay
Interest	.001*	.99	.06
Pressure	.26	.13	.45
Choice	.09	.04*	.55
Value	.24	.40	.18
Effort	.96	.33	.37
Competence	.98	.88	.70

Note. * $p < 0.05$

Table 4.17: Means and SD of Engagement for Participation and Gameplay conditions

Participation		Gameplay					
		VG n = 38		MG n = 39		Total n = 77	
		M	SD	M	SD	M	SD
		Group VPVG n = 12		**Group VPMG** n = 13		**VP** n = 25	
VP n=25	Engagement	7.3	1.4	7.9	1.2	7.6	1.3
		Group MPVG n = 26		**Group MPMG** n = 26		**MP** n = 52	
MP n=52	Engagement	6.9	1.5	6.7	1.8	6.8	1.6
		VG n = 38		**MG** n = 39		**Total** n = 77	
Total n=77	Engagement	7.0	1.5	7.1	1.7	7.0	1.6

4.4.6   Discussion

The outcomes of Experiment 3 are in line with to those of the previous experiments. Experiment 3 shows that the outcomes of the previous experiments hold up in a more formal and more relevant setting. Mandatory participants (MP) and players (MG) do not report significant negative effects of the obligation.

There were new findings with regard to the age of the participants and technical difficulties. Below, we will discuss (A) learning effect, (B) gameplay experience, (C) motivation, (D) non-gamers, (E) incentives and rewards, (F) cultural aspects, (G) age, and (H) technical difficulties.

A. Learning effect

The third experiment in the study of voluntary play was aimed to determine the effect of voluntariness with mandatory participation (MP) in a formal setting. After Experiment 1 and 2, we concluded that the *CloudAtlas* game has no learning effect and that time spent on training is not a factor. In Experiment 3, we found no significant difference in test scores between the four test groups. The participants in the MG groups played longer than the VG participants but did not achieve a higher test score.

The participants in the third experiment did achieve higher test scores than participants in the previous experiments. This may be related to the background of the participants. All participants of Experiment 3 were involved in aviation and flight training, whereas in the previous experiments only a few individuals were.

The lack of learning effect from the game most likely indicates issues with the game or the test. The game is not as effective as intended, or the test is not valid.

B. Gameplay experience

Only three participants deliberately chose not to play the game, but again, most VG participants ended their gameplay within the advised ten minutes, that was also the amount of mandated gameplay for the MG participants.

The reasons why VG participants stop playing so soon remains unknown. They report a similar enjoyment of the game and a similar engagement of the entire training. Almost two-thirds of the MG participants play more than two minutes beyond the ten-minute minimum, showing that the game, in fact, can be engaging. These outcomes may again indicate that a minimum time requirement is beneficial for gameplay, as it forces the participant not to give up at the first setback.

C. Motivation

A significant effect of the Participation group was found on motivation. This is in accordance with our expectation that volunteers are more motivated than participants who do so out of obligation. In the VP group, a participant who did not feel attracted to the study could decide not to participate. In the MP group, a participant with a similar feeling had to participate and hence may have reported a lower motivation. This is also supported by the outcome that VP participants score higher on the IMI subscale of Interest. We have found that VP participants feel more positive about being obliged to play the game for a minimum amount of time and that they report a higher engagement in the entire training.

VG participants reported that they experienced more freedom in the experiment, but they do not rate the game or the training different from the MG participants.

D. Non-gamers

In Experiment 3 we did not focus on differences between gamers and non-gamers, as the results of Experiment 1 showed that this did not have an effect on our measurements.

E. Incentives and rewards

In Experiment 3, no incentives or rewards were offered to any of the participants.

F. Cultural aspects

In Experiment 3, all participants were recruited within the Netherlands.

G. Age

Participants in Group VPMG and Group VPVG are significantly older than the participants in Group MPMG and Group MPVG. This is inherent to the way the participants were recruited. The MP participants are currently enrolled in flight schools to obtain their pilot licences. Most of them are in their early twenties. The VP participants have been invited to participate through newsletters and forum messages targeted at pilots who may have had their licences for quite some time.

Within the VP group, the participants in Group VPVG are significantly older than those in Group VPMG. As the assignment to the gameplay groups was done before starting the experiment, this is a random effect.

Age is strongly related to several outcomes (Table 4.18). Older participants achieve lower game scores and do not play as long as younger participants. They report to experience a higher engagement, higher interest and a more positive feeling about the obligation, but they also experience more pressure. Despite these correlations, ANCOVA testing shows that age does not have a significant effect on the outcomes of the analyses.

H. Technical difficulties

The technology used to develop the game limited its use to specific internet browsers on desktop and notebook computers. Smartphones and tablets could not be used. Due to this restriction, six participants in the VG group reported technical difficulties. As MG participants with technical difficulties could not finish the experiment and share their remarks, they will most likely have abandoned the experiment altogether. As the participants who reported the technical difficulties were unable to play the game, they have been left out of the analyses. However, some participants may have run into these difficulties and switched to a different device or internet browser. This may have had an effect on the enjoyment of the game, and the overall appreciation of the experiment that we cannot discern.

Table 4.18: Outcomes significantly correlating with Age

Outcome	Correlation		
	N	r	p
Self-reported prior knowledge	77	0.23	<0.05
Highest game score	77	-0.37	<0.01
Time played in the game	77	-0.24	<0.05
Engagement in the training	77	0.24	<0.05
IMI subset Interest	77	0.28	<0.05
IMI subset Pressure	77	0.24	<0.05
Feeling about the mandatory time of gameplay (MG only)	39	0.39	<0.05

4.4.7   Section conclusion

Experiment 3 is in line with the outcomes of the previous experiments. Consistent with Experiments 1 and 2, we did not find a learning effect from playing the *CloudAtlas* game. Therefore, we cannot answer RQ 2a.

Despite a higher sense of freedom, voluntary players do not perform better in the game or on the test (Measurement 1) or play longer (Measurement 3). We were surprised to find that mandatory players (MG), and even mandatory participants (MP), enjoy the game as much as voluntary players and participants (Measurement 4). With regard to RQ 2b, we may conclude that the obligation does not take the fun out of the game; an interesting outcome.

4.5   Overall discussion of the experiments

Initially, we set out to show the effect of voluntariness on the learning effect of a serious game. The voluntary part turned out to be easy. However, we have established that it is very hard, perhaps impossible, to create a genuinely mandatory setting. This made it hard to compare voluntary gameplay with a type of gameplay that would resemble assigned coursework in an educational or vocational setting.

After having all volunteers in the first experiment, we tried to create a stronger distinction between voluntary and mandatory gameplay by adding voluntary (VP) and mandatory participation (MP) to the second experiment. This proved to be difficult. The Dutch educational system requires a teacher to publish all requirements of a higher education course. It does not allow adding additional requirements later on. As a result, it was difficult to enforce the students' participation because they still had a possibility not to participate. Hence, students were still somewhat free to decide whether to participate or not. The amount of obligation perceived depended strongly on the personality of the student.

In the third experiment, we tried again. We recruited our participants from Dutch flight schools. As these institutions fall outside the regular Dutch education system, we expected them to have more room to require participation from their students. Moreover, we expected aspirant pilots to have a certain respect for their flight instructors. The third experiment did have a higher percentage of participation in the mandatory participation (MP) group, but still, some students did not participate. Apparently, they did not experience the obligation as such.

Mandatory participation is not a dichotomous variable. Some of the mandatory participants (MP) felt they had volunteered, others did feel somewhat obliged, but did not have negative feelings about it. Also, there were students who did not participate, despite being assigned to do so by their instructors. None of our experiments had a situation of undeniable mandatory participation.

4.6   Limitations

In the experiments, we faced two types of limitations that may have influenced the results. First, the experiments had a small number of participants (Subsection 4.6.1). Second, we were unable to create a strong feeling of obligation in the mandatory participants (Subsection 4.6.2).

4.6.1   Number of participants

Experiment 1 had a small number of participants. By recruiting through social media, we aimed to reach a large number of participants, but in fact, the number of participants was small. The group difference on prior motivation would probably not have occurred with a larger sample size or a different assignment strategy (pair matching).

In Experiment 2, recruiting was done through institutions for higher education rather than social media. In Experiment 3, participants were recruited through flight schools and aviation-related organisations. This led to the recruitment of more participants, but overall, the number of participants was still small.

4.6.2   Mandatory participation

In Experiment 1, all participants were voluntary participants. There was no obligation to participate in the study. Mandatory participation in the study (as a part of a regular course) would be of interest as this would provide a normal motivation setting for students in which the effects of voluntary gameplay can be observed without self-selection issues.

In the design of Experiment 2 and 3, we attempted to counteract the limitations of Experiment 1. In Experiment 2 and 3, the self-selection issue was addressed by having participation assigned by teachers and flight instructors respectively. The assigned participation seems to have resulted in a stronger obligation, but it still was not as strong as we expect it to be in a setting of game-based learning.

4.7   Chapter conclusion

We conducted three experiments to measure the effect of voluntary play on the outcomes of a serious game. In our experiments, we were unable to create a strong sense of obligation through mandatory participation.

We will address the research question in Subsection 4.7.1. Next, we will discuss the implication of our outcomes in Subsection 4.7.2. Finally, we will look at future research in Subsection 4.7.3.

4.7.1   Answering research question 2 (RQ 2)

This explorative study aimed to answer RQ 2 and its subquestions.

RQ 2: *What is the effect of voluntary play on the outcomes of a serious game?*

RQ 2a: *To what extent does the voluntary play of a serious game affect the learning effect?*

RQ 2b: *To what extent does the voluntary play of a serious game affect the gameplay experience of the player?*

To answer RQ 2a, we look at the outcomes of the game, measured with the knowledge and application questions in the test in all three experiments. The *CloudAtlas* game was developed specifically for this study. The outcomes indicate that players may need some time to get engaged in the game. This may have to do with the quality, playability, difficulty, or even graphical design of the game. In an ideal situation, a serious game will lead to a distinct improvement of player performance in a test regardless of voluntary or mandatory use. Initially, we expected that using a serious game voluntarily would lead to a greater improvement, in comparison to mandatory use. However, the *CloudAtlas* game seems to have no learning effect at all. Playing the game does not make a difference in the test scores, regardless of whether the game was played voluntarily or mandatorily (Measurement 1).

To answer RQ 2b, we look at the enjoyment, engagement (Measurement 3) and the feelings about being obliged to play (Measurement 4). In all three experiments, mandatory players report a similar enjoyment or engagement as voluntary players. They do not have strong negative feelings about the obligation (Measurement 4). Moreover, they play longer than the voluntary players (Measurement 2).

In none of the three experiments, we found that voluntary play has an effect on the learning effect or the gameplay. We did not find evidence that the voluntariness of gaming, highly rated by many game theorists and practitioners, is essential in order for the game to be fun. On the contrary, our findings indicate that participants with a stronger obligation play as long and as well as participants who were free to play the game. Also, they enjoy the game equally. Furthermore, we found that a little coercion increases the time spent in the game, which in turn may improve the learning effect.

The surprising finding that mandatory gameplay in the game does not appear to ruin the enjoyment and engagement in the game challenges the assumption of many game design theorists and practitioners that games need to be played voluntarily in order to be engaging, fun, and effective. The findings in our three experiments indicate that mandatory playing of a serious game is just as much fun as playing it voluntarily.

4.7.2 The implications of the outcomes

As voluntary play is thought to be indispensable for gaming, and thus for GBL, using a serious game in a training curriculum would disadvantage students who do not wish to play the game. Prensky [160] believes that a choice needs to be offered when using games for learning, to keep the voluntary aspect of gaming intact. If not playing the serious game would mean that the student does not learn, an alternative should be available. An alternative method would have to be offered to give them a fair chance of successfully completing the course, leading to higher training expenses.

Our experiments indicate that mandatory play of a serious game does not negatively affect the learning outcome or the player experience. If the findings of our study hold up in continued research, it may be concluded that providing the alternative method is not necessary, as mandatory players are not negatively affected by the obligation of playing a game. The absence of a negative effect of mandatory gameplay may open doors for education and professional training to implement GBL in their curriculum.

4.7.3   Future research

Our study found no evidence that voluntary gameplay is necessary for GBL. We found that mandatory participation (MP) and mandatory gameplay (MG) did not have significant negative effects on the enjoyment of the game at hand.

However, we used a game that turned out not to have a learning effect. In future research, it would be of interest to look at the effects of voluntary and mandatory participation and gameplay on the outcomes of a serious game that has been validated and proven effective. This would allow a conclusion to be drawn about the effect of voluntary play on the actual learning outcomes of the serious game.

Furthermore, future research should answer the question whether the effect of voluntary gameplay on the enjoyment of a serious game remains negligible as the type of game, the players and the intended learning outcomes are varied.

Thus, to definitively answer the question about the effect of voluntary play, further research is needed in various areas. Still, it is our expectation that voluntary play will not play as big a role as some games experts believe it does.

4