Exploring the use of arcade game elements for attitude change: two studies in the aviation safety domain

Luca Chittaro*, Fabio Buttussi*

Human-Computer Interaction Lab, University of Udine, Italy

* Corresponding authors. Tel.: +39 0432 558450; fax: +39 0432 558499.

E-mail addresses: luca.chittaro@uniud.it (L. Chittaro), fabio.buttussi@uniud.it (F. Buttussi)

Abstract

Gamification and serious games are increasingly employed for attitude change purposes. However, they frequently exploit a limited set of game elements, such as points, badges and leaderboards, and scarcely explore the use of more complex elements that make games engaging. In this paper, we focus on game elements from the arcade genre, and their use in serious games for attitude change. In particular, we propose a serious game for mobile devices that employs arcade game elements for attitude change in aviation safety, and we assess its effects with two different studies. The first study compares the immediate effects of the serious game vs. the traditional approach (safety card), showing that the game is more engaging and can improve attitudes towards aircraft emergencies in terms of users' self-efficacy and perceived vulnerability to the risk. The second study assesses the effects of the serious game when used over time (1-week) in a naturalistic setting, showing that the game can engage users and increase knowledge about correct and wrong behaviors. In both studies, engagement turned out to play a role that we analyze in detail.

Keywords

Serious games; arcade genre; attitude change; engagement; user study; aviation safety

1. Introduction

Gamification can be defined as the incorporation of game elements into interactive applications without a full-fledged game as the end product (Deterding et al., 2011; Deterding, 2012; Seaborn and Fels, 2015), while serious games (SGs) are (full-fledged) games that use entertainment to further training, education, health, public policy, and strategic communication objectives (Zyda, 2005). Over the last years, gamification and SGs have been successfully employed in several domains (Connolly et al., 2012; Hamari et al., 2014; Laamarti et al., 2014; Seaborn and Fels, 2015). In some cases, the proposed design was aimed at attitude change, i.e. a modification of "a person's overall evaluation of persons (including oneself), objects, and issues" (Petty and Wegener, 1998). For example, some gamification and SG proposals foster attitude change towards a brand or service (Hamari, 2017; Wise et al., 2008), or towards health behaviors (Baranowski et al., 2008; DeSmet et al., 2014).

However, several gamified applications still employ only a limited set of game elements such as points, badges and leaderboards (Rapp, 2017), raising criticism especially from the game community. For example, Robertson (2010) considers these elements as the "least essential to games" (Robertson, 2010), and Bogost (2011)

criticizes the gamification approach claiming that it “confuses the magical magnetism of games for simplistic compulsion meted out toward extrinsic incentives”. Similar criticism can be extended also to SGs, which are often based on simulations of real-world scenarios that integrate points, badges and leaderboards to reward the users, e.g. (Backlund et al., 2007; Mayer et al., 2013; Zyda, 2005), while several different elements that make games engaging are scarcely explored in SGs. More generally, several researchers suggest going beyond current gamification practices, and experimenting with new game elements (Baranowski et al., 2008; DeSmet et al., 2014; Nicholson, 2012; Rapp, 2017; Sakamoto et al., 2012).

Among game elements that are scarcely explored, this paper focuses on those that characterize the arcade genre. Although arcade games vary greatly in the tasks that players must perform (e.g. shooting, racing, fighting, or finding their ways in a maze), most of them are characterized by these peculiar elements:

- They are typically organized in levels of increasing difficulty. The first levels are easy, and players can rapidly learn simple controls to succeed. Then, they rapidly become more challenging, and require more skills (e.g. hand-eye coordination) to be completed.
- The pace is typically fast because there is a limited amount of time to complete each level or there is an increasing number of obstacles or enemies to avoid or fight.
- Graphics are pleasant but simple, and the environments and characters are often depicted in a cartoon, colorful and fun style.

These game elements contributed to make the arcade genre engaging for different generations of users, from those who played on cabinets in the ‘70s to those currently playing arcade games on mobile devices or on the Web. Therefore, we decided to explore their use in SGs with the aim of engaging users as well as positively changing their attitudes in a given domain.

In particular, we propose a SG for mobile devices that employs arcade game elements to deal with the topic of correct evacuation behaviors in aircraft emergencies (e.g. rapidly finding the path to the closest safe exit; avoiding fire, smoke, and unruly passengers), and foster attitude change towards aircraft emergencies. Different aspects of attitude change can be assessed in this domain. First, attitude change can appear in terms of modifications in people’s level of agreement on correct or wrong behaviors. Since people can ignore which behaviors are correct or wrong, increasing knowledge about the topic can contribute to attitude change. Second, people’s attitude towards aircraft emergencies can be measured in terms of psychological constructs of control such as self-efficacy, i.e. the belief an individual has on his/her ability to execute a behavior, which significantly determines performance outcomes (Bandura, 1997; Bandura, 2001). Third, attitude change can also be influenced by how people perceive a risk. Two important indicators of risk perception are vulnerability, i.e. how personally susceptible a person feels to a threat, and severity, i.e. how severe a person believes that the threat would be to his/her own life. Fourth, a further aspect of people’s attitude towards a topic that can be assessed is motivation to learn more about the topic.

To thoroughly explore if and how the proposed SG based on arcade game elements can play a role in attitude change, we carried out two different user studies. The first study investigated the immediate effects of the SG by assessing if it could engage users and foster attitude change in terms of agreement on correct and wrong behaviors, self-efficacy, vulnerability, and severity. The study assessed these measures before and after users tried the SG, and compared the game with the approach traditionally used to convey the same information, i.e. the safety card employed by airlines. Finally, the study explored the relations between engagement and the other considered variables.

The second study assessed instead the effects of the SG over time in a naturalistic setting, with users free to employ it as much as they wanted over a 1-week period. The study had two main objectives: assessing if the game could increase users' knowledge about correct and wrong behaviors, and investigating in detail the possible role of engagement in determining attitude change towards aircraft emergencies, including users' motivation to learn more about the topic.

The paper is organized as follows. In Section 2, we illustrate the game elements used in related work concerning gamification and SGs for attitude change. Section 3 illustrates the SG we developed by using elements from the arcade genre to change attitudes towards aircraft emergencies. Section 4 and 5 describe, respectively, the first and the second study, reporting and discussing the results. Section 6 concludes the paper outlining future work.

2. Related work

Different surveys dealt with the effects of gamification and SGs (Baranowski et al., 2008; DeSmet et al., 2014; Hamari et al., 2014; Seaborn and Fels, 2015). Two of them specifically focused on attitude change in a particular domain, i.e. health (Baranowski et al., 2008; DeSmet et al., 2014). This section will deal with the game elements employed for attitude change in different domains. Table 1 lists the different proposals, indicating the followed approach (gamification or SG), the employed game elements, the domain, and the topic of the proposal. In the following, we briefly summarize the game elements used in different domains, providing more information on the proposals that are more relevant to our work.

In the health domain, different authors followed a gamification approach based on points, badges, and other rewards for attitude change purposes (Cafazzo et al., 2012; Consolvo et al., 2008; Rose et al., 2013; Stinson et al., 2013). Rewards included the possibility to redeem digital items (Cafazzo et al., 2012), graphical rewards to visually enrich a virtual garden (Consolvo et al., 2008), and videos with popular actors (Stinson et al., 2013).

Similar game elements were used also in various SGs that required users to perform physical activity. For example, Lin et al. (2006) rewarded teams of users who walked in the real world with the growth and activity of virtual fishes, and created a leaderboard showing the fishes of teams with highest scores to stimulate cooperation and competition. Some SGs used a story to convey health messages (see Baranowski et al., 2008). Unfortunately, only a few proposals explored additional game elements. Notably, game elements from the platform genre (e.g. running, jumping, and taking or avoiding items) were used in a SG (Fuchslocher, 2011) that required players to balance the blood glucose level of their character (e.g. by taking food or insulin items). The elements of two popular arcade games (Arkanoid and Snake) were instead used in two SGs controlled by users' physical activity (Buttussi et al., 2007; Chittaro and Sioni, 2012).

In their review of SGs to promote health-related attitude change, Baranowski et al. (2008) found overall positive effects, but also called for more research on the optimal use of game elements in the field. Similarly, the meta-analysis of DeSmet et al. (2014) found positive, but small effects of SGs on health-related attitude change, and encouraged further explorations to understand which game elements could create larger effects.

In the environmental sustainability domain, the different proposals resorted mainly to the common gamification approach based on points, rewards, and leaderboards (Gnauk et al., 2012; Liu et al., 2011; Massung et al., 2013). Interestingly, Liu et al. (2011) added graphical negative feedback by raising the water level in a virtual island when users did not perform the correct behavior.

In the marketing domain, gamification and SGs have been used to change people's attitudes towards brands, products, and services. As described in Wise et al. (2008), the typical approaches used range from featuring a brand inside an existing, popular game to create SGs specifically around a product. The first approach, called in-game advertising, is similar to brand placement in television programs and movies (Yang et al., 2006), and makes the brands or products recede behind game rules (Mau et al., 2008). In the second approach, called advergames or ad-games, the brands or products take the center stage in a SG whose rules are structured around the advertising message (Mau et al., 2008). An alternative approach consists in the gamification of the service itself as proposed by Hamari (2017), who employed badges to increase trade proposals, transactions, and comments in a peer-to-peer marketplace.

In the domain of safety, only a few studies focused on the effects of gamification and SGs on attitude change. Two of these studies (Chittaro and Sioni, 2015; Chittaro, 2016) concerned SGs based on simulations of real world scenarios, which employed fearful realistic feedback (e.g. showing blood or broken bones) to scare users about the negative consequences of not adopting the proposed behaviors, as suggested by fear appeals models such as Protection Motivation Theory (Rogers, 1983). An alternative approach was proposed by Burrows and Blanton (2016), who embedded graphical advertising about safe driving in the background of the scenes of a first-person shooter. Notably, they found a correlation between engagement in the game and the obtained players' positive attitudes towards correct behaviors.

In summary, game elements such as points, badges, and other forms of rewards were used for attitude change often and across domains in both gamification and SG proposals, while other game elements were only rarely considered. Different authors (Baranowski et al., 2008; DeSmet et al., 2014) pointed out the need to explore a wider set of game elements and to assess their effects on attitude change. This is particularly important in the safety domain where the few proposals that assessed attitude change mostly relied on simulations of real world scenarios that followed a fear-based approach. Therefore, the present paper focuses on different game elements, from the arcade genre, and studies their use in a SG for safety-related attitude change that does not follow a fear-based approach.

Table 1: Gamification and serious game proposals for attitude change in different domains.

Reference	**Approach**	**Game elements**	**Domain**	**Topic**
Burrows and Blanton (2016)	Serious game	Game elements from first-person shooters	Safety	Safe driving
Buttussi et al. (2007)	Serious game	Game elements from Arkanoid arcade game	Health	Engagement in a physical activity
Cafazzo et al. (2012)	Gamification	Points, rewards	Health	Measurement of blood glucose level in diabetic children
Chittaro and Sioni (2012)	Serious game	Game elements from Snake arcade game	Health	Engagement in walking
Chittaro and Sioni (2015)	Serious game	Simulation of a real world scenario	Safety	Terror attack preparedness
Chittaro (2016)	Serious game	Simulation of a real world scenario	Safety	Brace position in aircraft emergency landings
Consolvo et al. (2008)	Gamification	Graphical rewards	Health	Engagement in physical activities

Fuchslocher (2011)	Serious game	Game elements from platform genre	Health	Balance of blood glucose level in diabetic people
Gnauk et al. (2012)	Gamification	Points, leaderboards	Environmental sustainability	Engagement in an energy management system
Hamari (2017)	Gamification	Badges	Marketing	Engagement in a peer-to-peer marketplace
Lin et al. (2006)	Serious game	Rewards, leaderboard	Health	Engagement in walking
Liu et al. (2011)	Gamification	Points, graphical negative feedback	Environmental sustainability	Sustainable behaviors
Massung et al. (2013)	Gamification	Points, rewards	Environmental sustainability	Engagement in pro-environmental activities
Rose et al. (2013)	Gamification	Points	Health	Measurement of blood glucose level in diabetic people
Stinson et al. (2013)	Gamification	Badges, rewards	Health	Pain diaries of young cancer patients

3. The proposed serious game

The safety topic considered by our game is the emergency evacuation of an aircraft. Fast and safe evacuation during emergencies is fundamental, because the cabin becomes unsurvivable in about two minutes since fire erupts (Miur, 2004). Unfortunately, incident and accident reports describe a wide range of inappropriate behaviors performed by passengers during emergency evacuations that jeopardize their and others' survival (National Transportation Safety Board, 2000). The aim of the SG is to change users' attitudes towards aircraft emergencies and to promote the correct behaviors passengers should follow to evacuate safely. More specifically, the correct behaviors that the SG aims at promoting concern the need to: i) locate the closest exit, and use it as a first choice in the evacuation, ii) find an alternate exit when the closest one is not accessible, iii) leave luggage on the aircraft, iv) avoid smoke inhalation, v) stay away from fire, vi) be careful about unruly, competitive behaviors who disrupt the evacuation such as pushing.

Previous research simulated evacuation scenarios in their dramatic and fearful aspects (Chittaro and Buttussi, 2015; Buttussi and Chittaro, 2018), but the fear-based approach may not be attractive for some users. Therefore, the present paper explores a different approach. The aim is to turn possibly scary evacuation scenarios into fun, humorous situations with cartoon characters and objects, while at the same time keeping the safety messages clear: the SG can be successfully completed only by performing the correct safety behaviors. Different safety messages, such leaving a place within a given amount of time or avoiding dangerous areas and unruly people, recall typical game elements from the arcade genre, such the limited amount of time and the avoidance of obstacles and enemies. Therefore, we propose a SG for mobile devices that uses arcade game elements to promote the correct behaviors for a successful evacuation.

More precisely, the message about the importance of a fast evacuation is conveyed by giving time pressure a fundamental role in the game. As shown in Figure 1, a prominent clock with a countdown in seconds is displayed at the top of the screen. If the countdown completes before the player reaches an exit, then the goal is not achieved, and the level must be played again. The next level is unlocked only when the current one is completed successfully. Each level is short due to time pressure, and the difficulty increases level after level, as it typically happens in most arcade games.

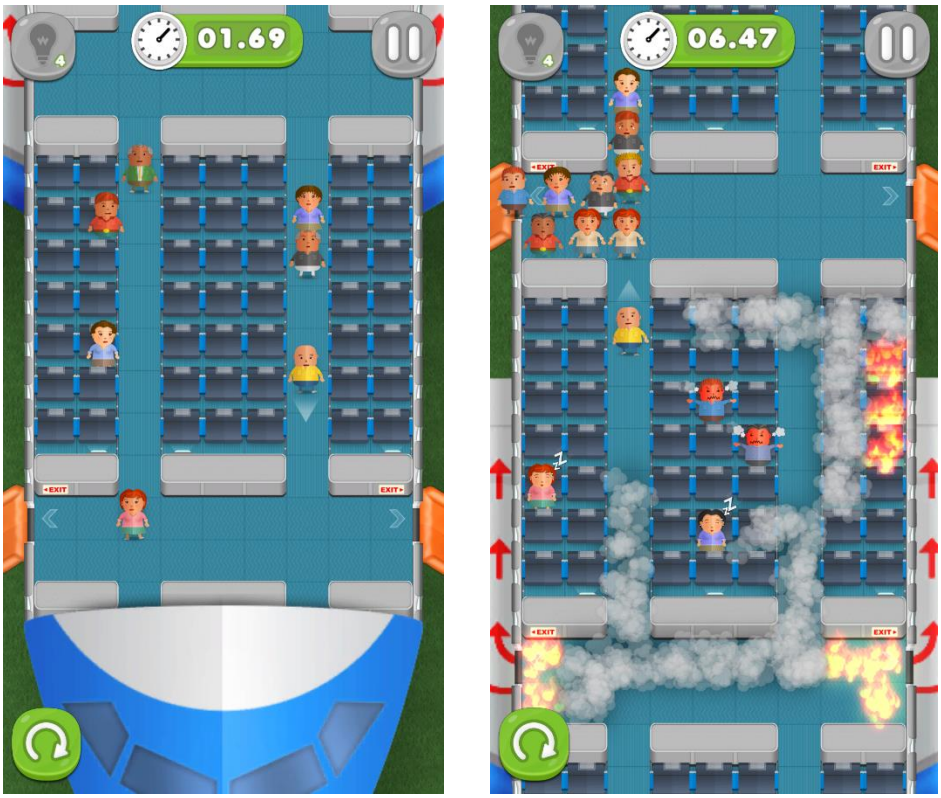

Figure 1. One of the easiest and one of the most difficult game levels.

The player is influenced into locating the best exit for the depicted circumstances as follows. In the easiest levels, all exits are accessible, but the available time allows the player to reach only the closest one. In the more difficult levels, an increasing number of exits becomes inaccessible due to smoke or fire, requiring more complex reasoning to reach the best exit for an evacuation that can be completed in the available time.

The message about luggage is conveyed through the behavior of other passengers, with some of them wrongly taking luggage (Figure 2). In the first levels, all passengers leave luggage on the plane, and move fast. As the level number increases, some passengers start to carry luggage (also in the form of unexpected objects like a TV set, a large beach ball, a fish bowl,...), and slow down the evacuation of the other characters (creating lines of people) and of the player. This makes it more difficult to find a path that leads to an exit in time. We exploited the same idea for the message about unruly behaviors. In the first levels, all passengers behave in an orderly manner. As the level of difficulty increases, some passengers will start to look angry (Figure 2), and act disorderly, making the evacuation chaotic. If the player gets into contact with them, they will push him/her on the ground, ending the level.

Players can control their character in real-time by dragging the finger anywhere on the screen in the desired direction of movement as if they had a virtual arcade joystick. As it is typical in arcade games, we kept the controls deliberately simple, so that players can rapidly learn them and easily succeed in the first levels. More skills are instead required to complete the more difficult levels successfully.

Finally, simplicity of graphics (Figure 1) and use of fun cartoon characters (Figure 2), which are characteristic of arcade games, were adopted too for the proposed game.

The game was implemented in C# using the Unity 5.3 game engine. We implemented two versions of the game. The first had 12 levels and was used in the study to assess the immediate effects of the SG, while the second has 48 levels and was used in the study that assesses the effects of the SG when used over time. The differences between the two versions were the total number of the levels and the increase in difficulty between a level and the next one.

Figure 2. Examples of some of the cartoon characters in the game.

4. Study 1

To evaluate the immediate effects of the SG, we carried out a between-groups study. Half participants (Serious Game group) played the SG described in Section 3, while the other half (Safety Card group) examined a safety card that presented the same safety information.

4.1 Hypotheses

We formulated the following hypotheses:

1. Participants in the Serious Game group should be more engaged than those in the Safety Card group, since it is known that passengers tend not to pay attention to safety cards (Corbett et al., 2008), while a game can potentially be more attractive.
2. If the previous hypothesis holds, the SG might increase participants' level of agreement on proposed behaviors more than the safety card. More engaged users were more oriented towards correct behaviors in the study of a first-person shooter game that contained safety-related advertising (Burrows and Blanton, 2016), so it is interesting to assess if the same effect could hold in the different case of an arcade SG concerning the intended safety context and behaviors.
3. The possible increase in self-efficacy should be higher in the Serious Game group, since gaining experience in performing a given behavior is a major factor that contributes to increase self-efficacy (Bandura, 1997), and participants in the Serious Game group actively put the correct behaviors in practice to advance in the game.

The study was explorative with respect to vulnerability and severity, since previous proposals that assessed these variables employed fear appeals (Chittaro and Sioni, 2015; Chittaro, 2016), which are likely to affect risk perception, while the present SG follows a humorous approach.

4.2 Materials

Participants in the Serious Game group played the first 12-level version of the SG on an Android smartphone. Participants in the Safety Card group examined an A4-sized safety card, printed in color. The pictorials in the safety card were those employed by one of the largest world airlines. The safety card used in the study included only the pictorials that provide the same safety information provided by the SG.

4.3 Participants

The evaluation involved a sample of 40 participants (20M, 20F). Participants were volunteers who received no compensation and were recruited through personal contact. Age ranged from 18 to 63 (M=29.98, SD=12.82).

We asked participants to rate their frequency of use of video games on a 7-point scale (1=never, 2=less than once a month, 3=about once a month, 4=several times a month, 5=several times a week, 6=every day for less than an hour, 7=every day for more than one hour). Answers ranged from 1 to 7 (median=2.5; 15 participants never played video games, 5 played less than once a month, 2 played about once a month, 4 played several times a month, 7 played several times a week, 1 played every day for less than an hour, and 6 played every day for more than an hour).

We also assessed individual differences in frequency of air travel by asking participants to count their number of flights in the last two years, as in (Corbett et al., 2008). Each flight had to be counted individually (e.g., a round trip from airport A to airport C via a connection through airport B results in four flights). Answers ranged from 0 to 16 (M=2.65, SD=3.50).

Participants were assigned to the two groups in such a way that each group had 20 participants (10M, 10F each), and the two groups were similar in terms of age, frequency of video game use, and frequency of air travel. Lack of significant differences between the two groups was confirmed by independent samples t-tests for age and frequency of air travel, and by Mann-Whitney U test (used because the variable was ordinal) for frequency of video game use.

4.4 Measures

4.4.1 Engagement

To measure the engagement experienced by participants, we administered a questionnaire that asked participants to rate their level of agreement about six statements on a 7-point scale (1=not at all, 7=very). The six statements were: i) it was boring, ii) it was engaging, iii) it aroused emotions in me, iv) I would have liked to continue using it, v) while I use it, I forgot about the real environment where I was, and vi) it was fun. After inverting the scale of the first statement, the six ratings were averaged to form a reliable scale (Cronbach's alpha=0.79).

4.4.2 Behavior agreement

To measure participants' level of agreement on correct and wrong behaviors, we used a questionnaire with a list of behaviors that one might follow during the evacuation of an aircraft, and we asked participants to rate their agreement about the correctness of each behavior on a 7-point Likert scale (1=strong disagreement, 7=strong agreement). The list of behaviors included nine correct behaviors presented in the experimental condition, and three wrong behaviors that were the opposite of correct presented behavior. After inverting the scale of wrong behaviors, the ratings were averaged to form the "behavior agreement" measure. To measure changes due to the experimental condition, we administered the questionnaire twice: before (pre-test) and after (post-test) the experimental condition. Mean pre-test behavior agreement was 6.12 (SD=0.67). An independent samples t-test showed no significant difference in the pre-test scores between the two groups.

4.4.3 Self-Efficacy

To measure self-efficacy, we administered a questionnaire that adapted items from well-known self-efficacy questionnaires (Schwarzer and Jerusalem, 2015) to the considered aviation safety context. The questionnaire had six items: i) I feel able to deal with an emergency evacuation of an aircraft, ii) I would be able to deal with an emergency evacuation even if the aircraft is on fire, iii) I would be able to deal with an emergency evacuation even if one or more exits are blocked, iv) I would be able to deal with an emergency evacuation even if most of the passengers scream or cry, v) I feel confident of my ability to exit from the aircraft in time, and vi) I would be able to help passengers in need. Participants were asked to rate each item a 7-point scale (1=not at all, 7=very). We administered the self-efficacy questionnaire twice: before (pre-test) and after (post-test) the experimental condition. Answers were averaged to form a reliable scale (Cronbach's alpha pre-test=0.94, post-test=0.96). Mean pre-test self-efficacy was 3.31 (SD=1.76). An independent samples t-test showed no significant differences in pre-test self-efficacy between the two groups.

4.4.4 Vulnerability

To measure vulnerability, we administered the questionnaire used in (Chittaro, 2016), which in turn was adapted from (De Hoog et al., 2008). Participants were asked to think about an emergency landing and evacuation of an aircraft, and to rate three items on a 7-point scale (1=not at all, 7=very). The three items asked how vulnerable respondents perceived themselves to be with respect to such risk; how high they thought their risk of being involved in such situation was; and how high the probability of suffering personal negative consequences from such situation was. The questionnaire was administered twice: before (pre-test) and after (post-test) the experimental condition. Cronbach's alpha was 0.64 in the pre-test and 0.67 in the post-test. Mean pre-test vulnerability was 3.52 (SD=1.16). An independent samples t-test showed no significant differences in pre-test vulnerability between the two groups.

4.4.5 Severity

Severity was measured by means of the questionnaire used in (Chittaro, 2016), which in turn was adapted from (De Hoog et al., 2008). The questionnaire asked participants to think about an emergency landing and evacuation of an aircraft and to rate three items on a 7-points scale (1=not at all, 7=very). The three items asked how severe, harmful, and serious the consequences of such situation would be. The questionnaire was administered twice: pre-test and post-test. Cronbach's alpha was 0.82 in the pre-test and 0.86 in the post-test. Mean pre-test severity was 5.76 (SD=1.00). An independent samples t-test showed no significant differences in pre-test severity between the two groups.

4.5 Procedure

The experimenter contacted participants and asked them the initial demographic information (gender, age, frequency of game use, and frequency of air travel). Before starting the test, participants were kindly asked to switch off their phones or turn off ringtones and vibration. Participants in the Serious Game group (respectively the Safety Card group) were told that we were testing a mobile game (respectively a safety card). All participants were informed that they would be asked to fill a few anonymized questionnaires and that they would be interviewed about their personal opinions. Finally, participants were told that they could refrain from continuing the experiment at any time, without providing a reason to the experimenter. After participants gave their consent, they filled the pre-test questionnaire to measure their agreement on correct and wrong behaviors as well as the pre-test questionnaires assessing self-efficacy, vulnerability, and severity.

Participants in the Serious Game group were told that the game was made of 12 levels, and that there was no time limit to complete them. Similarly, participants in the Safety Card group were told that they could examine the safety card how long they wanted until they thought they had understood the pictorials. Participants in both groups were invited to ask for any clarifications before starting the trial, since they could not ask questions to the experimenter during the trial.

After the experimental condition, participants filled the engagement questionnaire as well as the post-test questionnaires assessing self-efficacy, vulnerability, and severity. Then, participants filled the post-test questionnaire to measure their agreement on correct and wrong behaviors.

4.6 Results

To compare the engagement of the two groups we used an independent samples t-test. For all the other dependent variables, which were measured two times each, we used a 2 x 2 mixed design ANOVA, in which group served as the between-subjects variable, and time of measurement (pre-test, post-test) served as the within-subjects variable. In case of interaction between group and time of measurement, we proceeded with the analysis of simple main effects as described by (Cohen, 2001). To test the effects of time of measurement separately for each group, we used one-way repeated measures ANOVA. To test the effects of group separately at each time of measurements, we used between-subjects ANOVA. Finally, to test the relations between engagement and the other assessed variables, we used Pearson's correlation.

4.6.1 Engagement

The analysis revealed that the difference in engagement between the two groups (Figure 3) was statistically significant, $t(38)=-3.90$, $p<0.001$, two-tailed. The SG was more engaging ($M=4.28$, $SD=1.18$) than the safety card ($M=2.99$, $SD=0.87$). The effect size was large (Cohen's $d=1.24$).

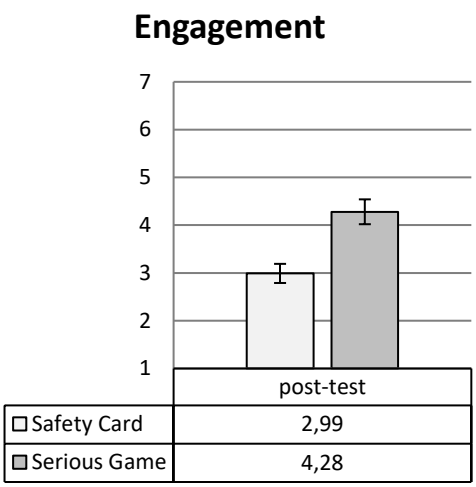

Figure 3: Engagement in the Safety Card and in the Serious Game group. Capped vertical bars indicate ± SE.

4.6.2 Behavior agreement

For behavior agreement (Figure 4), the analysis revealed no main effect of group, no main effect of time of measurement, and no interaction.

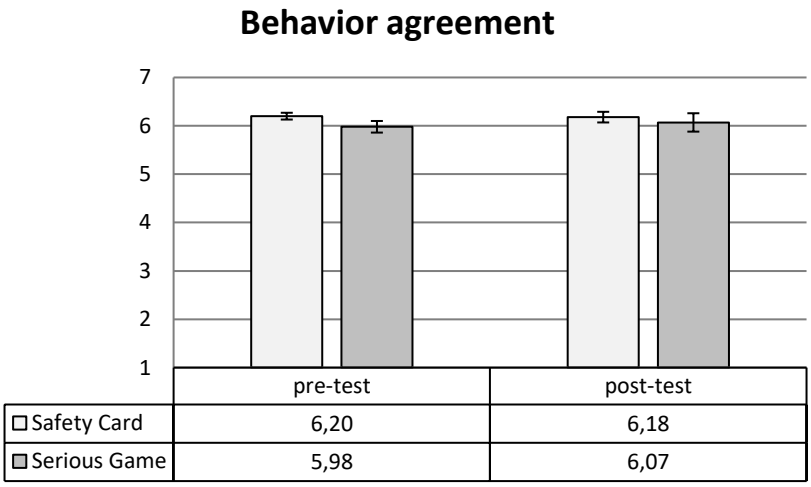

Figure 4: Behavior agreement in the Safety Card and in the Serious Game group before and after the experimental condition. Capped vertical bars indicate ± SE.

4.6.3 Self-Efficacy

For self-efficacy (Figure 5), the analysis revealed a main effect of group, $F(1,38)=6.10$, $p<0.05$, $\eta_p^2=0.14$, a main effect of time of measurement $F(1,38)=29.42$, $p<0.001$, $\eta_p^2=0.44$, and a group by time of measurement interaction, $F(1,38)=5.82$, $p<0.05$, $\eta_p^2=0.13$. We thus proceeded with the analysis of simple main effects. In the Safety Card group, the mean pre-test self-efficacy score was 3.00 (SD=1.29), and the mean post-test score was 3.48 (SD=1.34). The analysis revealed a statistically significant difference, $F(1,19)=12.69$, $p<0.01$, $\eta_p^2=0.40$. In the Serious Game group, the mean pre-test self-efficacy score was 3.63 (SD=1.58), and the mean post-test score was 4.88 (SD=1.34). The analysis revealed a statistically significant difference, $F(1,19)=18.69$, $p<0.001$, $\eta_p^2=0.50$. There were no statistically significant differences between the two groups at pre-test time, while we found a statistically significant difference at post-test time, $F(1,38)=10.90$, $p<0.01$, $\eta_p^2=0.22$, with a higher self-efficacy in the Serious Game group (M=4.88, SD=1.34) than in the Safety Card group (M=3.48, SD=1.34).

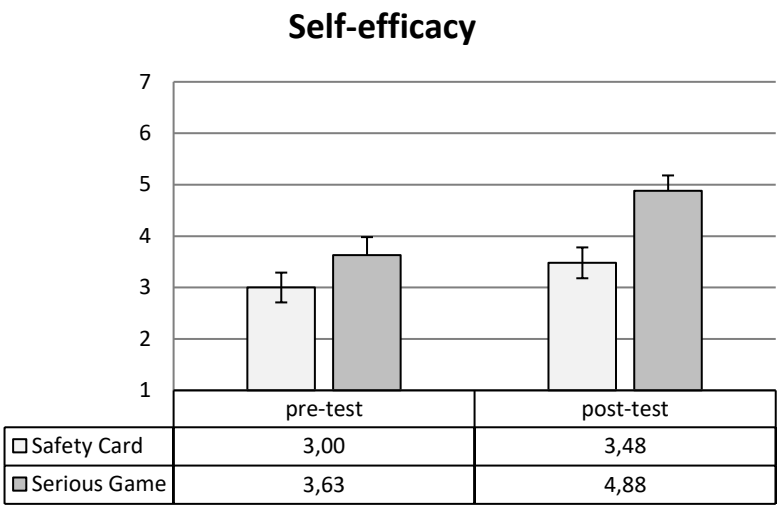

Figure 5: Self-efficacy in the Safety Card and in the Serious Game group before and after the experimental condition. Capped vertical bars indicate ± SE.

4.6.4 Vulnerability

For vulnerability (Figure 6), the analysis revealed no main effect of group, a main effect of time of measurement $F(1,38)=4.90$, $p<0.05$, $\eta_p^2=0.11$, and a group by time of measurement interaction, $F(1,38)=7.66$, $p<0.01$, $\eta_p^2=0.17$. We thus proceeded with the analysis of simple main effects. In the Safety Card group, the mean pre-test vulnerability score was 3.57 (SD=1.15), and the mean post-test score was 3.50 (SD=1.16). The difference was not statistically significant. In the Serious Game group, the mean pre-test vulnerability score was 3.48 (SD=1.20), and the mean post-test score was 4.08 (SD=1.21). The analysis revealed a statistically significant difference, $F(1,19)=9.11$, $p<0.01$, $\eta_p^2=0.32$. There were no statistically significant differences between the two groups at pre-test or post-test time.

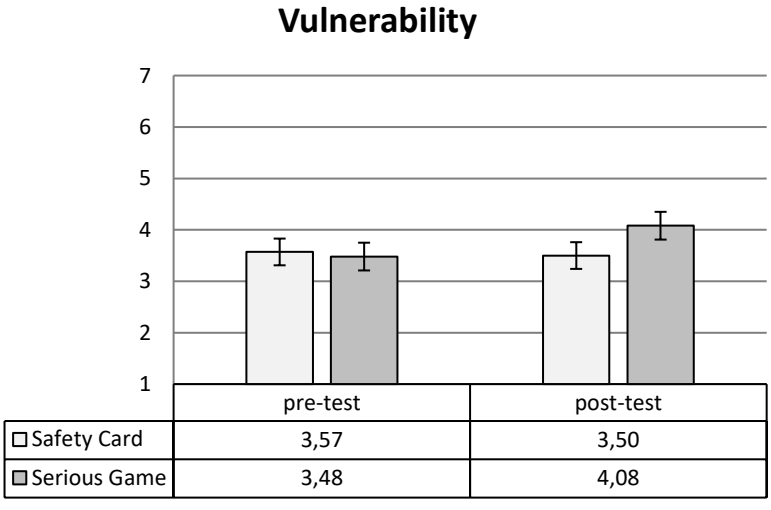

Figure 6: Vulnerability in the Safety Card and in the Serious Game group before and after the experimental condition. Capped vertical bars indicate ± SE.

4.6.5 Severity

For severity (Figure 7), the analysis revealed no main effect of group, no main effect of time of measurement, and no interaction.

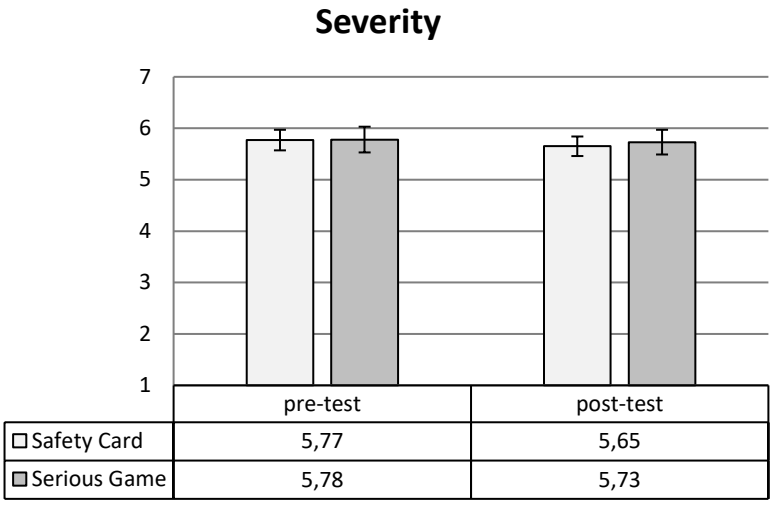

Figure 7: Severity in the Safety Card and in the Serious Game group before and after the experimental condition. Capped vertical bars indicate ± SE.

4.6.6 Engagement correlations

We found a statistically significant positive correlation between engagement and post-pre difference in behavior agreement, $\rho=0.48$, $p<0.01$. Also the positive relation between engagement and post-pre difference in self-efficacy was statistically significant, $\rho=0.33$, $p<0.05$. No significant relation was found between engagement and post-pre difference in vulnerability and severity. Analyzing engagement correlations separately for each group, we found that the relation with post-pre difference in behavior agreement for the Serious Game group was significant, $\rho=0.65$, $p<0.01$, while it was not for the Safety Card group. The relation with post-pre difference in self-efficacy did not reach significance for either of the two groups.

4.7 Discussion

Results about engagement confirmed our hypothesis: participants in the Serious Game group were significantly more engaged than those in the Safety Card group. Similar results were found in studies that compared safety cards with SGs based on fear appeals (Chittaro and Buttussi, 2015), but to the best of our knowledge, this is the first study that extends the exploration to the arcade genre. The use of arcade game elements, which made several entertainment games engaging, had likely contributed to the results, and could thus be considered in solutions to overcome passengers' lack of interest towards traditional safety media.

Unfortunately, the study did not show the expected effects of group and time of measurement on behavior agreement. Indeed, the analysis did not reveal any improvement on behavior agreement with either the game or the safety card. However, it must be noted that behavior agreement was already very high in the pre-test (6.12 on a 1 to 7 scale), so the lack of significant effects is partially explained by a ceiling effect. This ceiling effect could be due to different factors. First, most participants might have already known the considered safety information (previous knowledge). Second, they might have not known it, but felt to agree with correct behaviors after reading them in the initial questionnaire (questionnaire bias). The following facts are against the first factor: i) 14 of the 40 participants did not fly in the previous two years, so it is unlikely they were exposed to aviation safety information, ii) passengers tend not to pay attention to traditional safety materials and even those who pay attention show an unacceptable level of comprehension (Corbett et al., 2008), so even participants who had flown might not have been familiar with the relevant safety information. In any case,

since previous knowledge and/or questionnaire bias affected behavior agreement before trying the experimental condition, this first study highlighted that further research was needed to understand if the proposed approach could deliver the safety information required to support attitude change towards aircraft emergencies. This was one of the motivations for the second study, described in Section 5, which will address this issue.

This first study provided instead evidence to support our hypothesis about self-efficacy. The analysis showed a statistically significant increase in self-efficacy after the experimental condition in both groups, so both the SG and the safety card could successfully improve this aspect of participants' attitudes. Moreover, the difference in post-test self-efficacy between the two groups was statistically significant and indicated that the SG was more effective than the safety card in improving self-efficacy. These results support the adoption of SGs with arcade game elements in the safety domain: the application of engaging game elements had likely contributed to make players practice the correct behaviors to advance in the game, and even if the behaviors were practiced in a virtual, cartoon-like, and humorous context, the gained experience contributed to increase self-efficacy towards aircraft emergencies.

Of the two risk perception indicators that we explored, we found an interesting result for vulnerability. While in the Safety Card group the difference between pre-test and post-test was not significant, in the Serious Game group there was a significant increase in vulnerability after trying the experimental condition. This is likely due to the fact that, if participants did not follow the correct behaviors in the game, their character did not evacuate the aircraft in time, and the time pressure game element could have contributed to make participants more aware of their vulnerability. Notably, Cronbach's alphas were relatively low for this questionnaire. They were low also in a previous study which used the same questionnaire (Chittaro, 2016), while they were high in a study that used the questionnaire from which the employed one was adapted (De Hoog et al., 2008). In our case, we found that the third item (i.e., the one about the probability of suffering negative consequences from the situation) had the lowest correlation in the scale. Indeed, its values were higher than those of the other items even before the experimental condition, and their differences between pre-test and post-test were not statistically significant. The lack of significant changes in the values of this item could be due to the fact that the safety card did not show the negative consequences of not following the behaviors, and the game showed them in a humorous way, as in many arcade games, avoiding the use of fear-inducing cues. The same observation could explain the lack of significant differences for both conditions in the means of severity, which were very similar between pre-test and post-test. In addition, the means of severity were high even before the experimental condition, and thus the lack of post-pre differences could also be partially explained by a ceiling effect.

Finally, the study showed a positive correlation between engagement and post-pre difference in behavior agreement. More precisely, the higher was the engagement, the higher was the increase in behavior agreement, and this result was found despite the high pre-test scores in behavior agreement and the lack of significant differences between pre-test and post-test. Notably, the correlation was found when we considered all the participants, and when we considered only the Serious Game group, but not when we considered only the Safety Card group. A relation between engagement and agreement on correct behaviors was previously found for a game with safety-related in-game advertising (Burrows and Blanton, 2016). Unlike that game, which was a first-person shooter completely unrelated to the safety message contained in it, our study extended the finding to a SG that concerned the intended safety behaviors, and approached them in a humorous way by using game elements typical of arcade games.

Overall, this first study showed that the SG was more engaging than the traditional approach based on a safety card, and that it was also more effective to increase self-efficacy and vulnerability. The higher was the engagement, the higher was the increase in behavior agreement, but significant post-pre differences in behavior agreement were not found probably due to a ceiling effect.

5. Study 2

The second study was exploratory in nature and focused on better understanding the effects of the SG with users who played it over a 1-week period in naturalistic settings.

5.1 Aim

The aim of this study was twofold:

1. Assessing if the proposed approach could effectively deliver safety information about correct behaviors, wrong behaviors, and dangers during the evacuation of an aircraft.
2. Broadly investigating the role of engagement by considering Csikszentmihalyi's concept of flow experience, i.e., a situation of complete absorption or engagement in an activity (Csikszentmihalyi, 1990), by administering the EGameFlow questionnaire (Fu et al., 2009).

5.1 Materials

Participants played the full 48-levels version of the SG on an Android smartphone.

5.2 Participants

The evaluation involved a sample of 38 participants (24M, 14F). Participants were volunteers who received no compensation, and were recruited through personal contact. Age ranged from 18 to 61 (M=29.87, SD=12.09).

We asked participants to rate their frequency of use of video games on the same 7-point scale used in the previous study. Answers ranged from 1 to 7 (median=3; 11 participants never played video games, 4 played less than once a month, 5 played about once a month, 7 played several times a month, 5 played several times a week, 2 played every day for less than an hour, and 4 played every day for more than an hour).

As in the previous study, we also asked participants to count their number of flights in the last two years. Answers ranged from 0 to 50 (M=6.05, SD=9.72).

5.3 Measures

5.3.1 Knowledge

We asked participants the four questions reported in Table 2. The questions were about the same aviation safety topics addressed in the behavior agreement questionnaire described in Section 4. However, in this second study we did not measure agreement on listed behaviors, but we orally asked questions allowing for open answers to avoid possible ceiling effects due to the questionnaire. Questions Q1, Q2, and Q4 admitted a single correct answer each, while to correctly answer question Q3 participants had to list a total of six dangers and wrong behaviors. We counted the number of correct answers provided, so the knowledge score could

range from 0 to 9. The experience of playing the game should have taught participants how to correctly answer the first three questions. It did not instead allow one to experientially learn the correct answer to the fourth question, because there was no “bend” command in the game and participants had to keep their character away from smoke, but we included the question to check if participants might be able to reason further about the situation. To measure changes due to the experimental condition, we asked the questions twice: before (pre-test) and after (post-test) the experimental condition. Mean pre-test knowledge score was 4.61 (SD=1.31).

5.3.2 EGameFlow

The EGameFlow questionnaire (Fu et al., 2009) assesses eight constructs concerning video games: i) concentration, i.e. how much the participant can remain concentrated on the game, ii) goal clarity, i.e. how much the goals of the game are clear, iii) feedback, i.e. quality of perceived feedback provided by the game, iv) challenge, i.e. how appropriate is the level of challenge and its increase as the participant progresses in the game, v) autonomy, i.e. sense of control on the game perceived by the participant, vi) immersion, i.e. how much the participant feels immersed in the game, vii) social interaction, i.e., how much the game becomes a mean to socially interact among participants, viii) knowledge improvement, how much the participant feels that the game helps in improving his/her knowledge. To assess these constructs, the questionnaire asks participants to rate a total of 42 items on a 7-point Likert scale (1=strong disagreement, 7=strong agreement) and averages the items for each construct to form reliable scales. Since our game had no social feature, we removed the items related to social interaction and administered participants a questionnaire with the remaining seven constructs for a total of 36 items. It is important to note that the first six constructs are present also in the GameFlow questionnaire (Sweetser and Wyeth, 2005), and are connected to Csikszentmihalyi’s concept of flow experience (Csikszentmihalyi, 1990), while knowledge improvement is a construct specifically added by (Fu et al., 2009) for games that can be used for learning. The last item of EGameFlow concerns an attitude, that is motivation to learn more (“I want to know more about the knowledge taught”).

5.3.3 Maximum level reached and playing time

The game automatically logged the number of the maximum level reached by participants, and the minutes they spent playing the game during the 1-week period.

5.4 Procedure

Data were collected from participants at two different times. The first time, the experimenter collected demographic information (gender, age, frequency of game use, and frequency of air travel) and assessed participants’ knowledge (pre-test). Then, the experimenter introduced the participants to the game, helped them installing it on their personal smartphone, and tried the first level together to check that participants understood how to play the game. Then, the experimenter told participants that they could play the game as much as they wanted during the next 7 days, and that they would be asked to answer some questions at the end of the 1-week period. After the week had passed, the experimenter re-assessed participants’ knowledge (post-test) to measure possible knowledge gain and administered the EGameFlow questionnaire.

5.5 Results

To compare knowledge scores before and after the 1-week period, we used one-way repeated measures ANOVA, with time of measurement (pre-test, post-test) as the within-subjects variable. Post-pre differences in

single answers were instead analyzed using Wilcoxon signed-rank test. Finally, to test if there were relations between the assessed variables, we used Pearson's correlation.

5.5.1 Knowledge

Before playing the SG, the mean knowledge score (Figure 8) was 4.61 (SD=1.31). After playing the SG, it was 6.89 (SD=1.56). The difference was statistically significant, $F(1,37)=65.91$, $p<0.001$, $\eta_p^2=0.64$. Table 2 shows the mean scores for each specific correct answer at the beginning and at the end of the week, and indicates which differences were statistically significant.

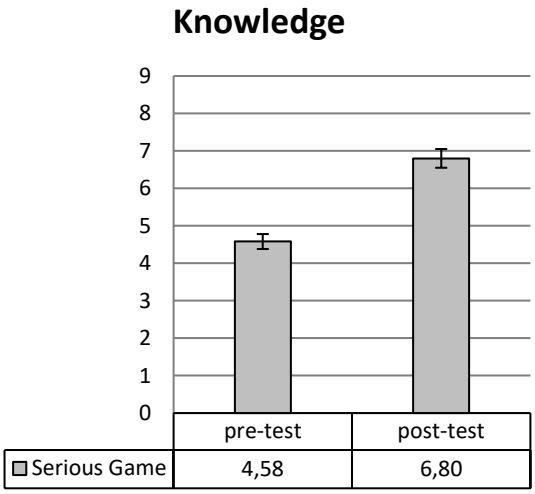

Figure 8: Knowledge before and after the week with the serious game. Capped vertical bars indicate ± SE.

Table 2: Oral questions used to assess participants' knowledge, correct answers (in italics), mean pre-test and post-test scores, and statistical significance of the post-pre difference.

Question and answers	**Pre-test**	**Post-test**	**Significance**
Q1: In the emergency evacuation of a plane, to which exit should a passenger go first? - *To the closest exit*	.50	.82	p=0.001
Q2: If the exit that the passengers have reached is not usable, what must they do? - *Locate and go to the closest usable exit*	.66	.97	p=0.001
Q3: What dangers can the passenger encounter during the evacuation? Which passenger's behavior can slow down the evacuation?			
- *Smoke*	.05	.68	p<0.001
- *Fire*	.21	.74	p<0.001
- *Unruly passengers*	.53	.61	not significant
- *Panic*	.92	.61	p=0.001
- *People taking luggage*	.13	.66	p<0.001
- *Crowds (moving slow, crowded exits)*	.29	.55	p<0.05
Q4: What should passengers do if they find themselves in smoke during the evacuation? - *Bend down to prevent smoke inhalation*	.39	.45	not significant

5.5.2 EGameFlow

The results obtained for the considered EGameFlow constructs were the following (Figure 9): i) concentration, M=5.58, SD=0.76, ii) goal clarity, M=6.26, SD=1.10, iii) feedback, M=6.04, SD=0.73, iv) challenge, M=5.11, SD=0.96, v) autonomy, M=4.77, SD=1.07, vi) immersion, M=3.45, SD=1.36, vii) knowledge improvement, M=5.14, SD=1.34.

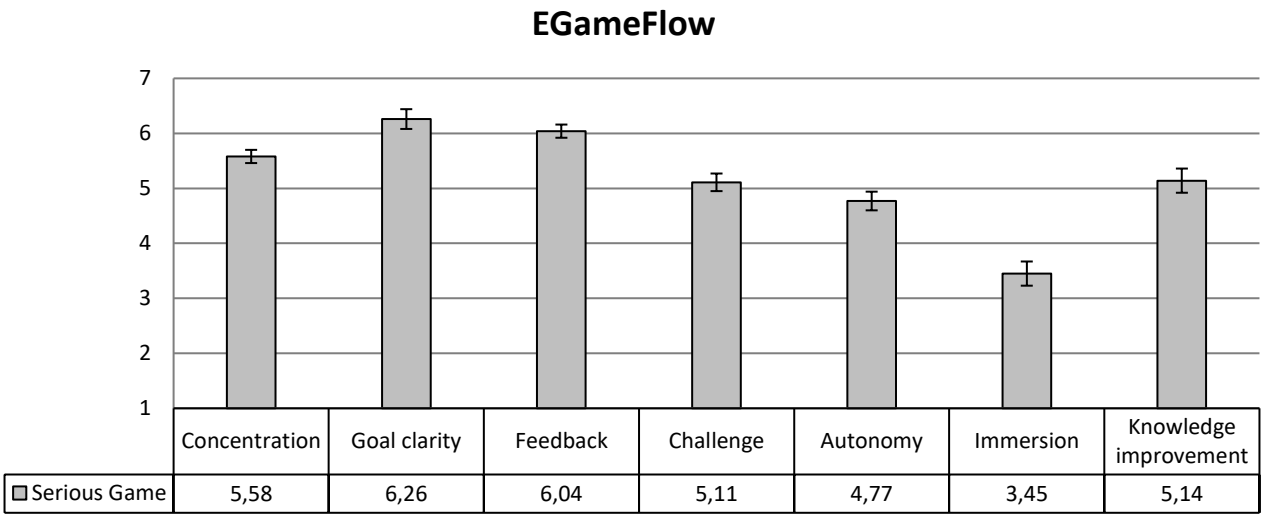

Figure 9: Participants' average ratings of the considered EGameFlow constructs after the week with the serious game. Capped vertical bars indicate ± SE.

5.5.3 Maximum level reached and playing time

Maximum level reached by participants ranged from 10 to 48 (M=41.61, SD=9.64). The level was low (10 and 14, respectively) only for two participants. The remaining 36 participants reached level 29 or higher, and 21 participants completed all 48 levels of the game. Playing time ranged from 11.83 to 159.62 minutes (M=47.39, SD=31.64).

5.5.4 Correlations

We examined possible correlations between EGameFlow knowledge improvement and the six constructs related to flow and engagement. We found a strong positive correlation between knowledge improvement and feedback, $\rho=0.59$, $p<0.001$. Knowledge improvement also positively correlated with autonomy, $\rho=0.44$, $p<0.01$, and immersion, $\rho=0.43$, $p<0.01$. Finally, some weaker correlations were found also with goal clarity, $\rho=0.38$, $p<0.05$, and concentration, $\rho=0.35$, $p<0.05$.

Examining in more detail the item concerning motivation to learn more, we found significant correlations with immersion, $\rho=0.48$, $p<0.01$, autonomy, $\rho=0.38$, $p<0.05$, and feedback, $\rho=0.35$, $p<0.05$. Moreover, motivation correlated also with participant's knowledge: a negative correlation was found with pre-test knowledge score, $\rho=-0.34$, $p<0.05$, and a positive correlation was found with knowledge gain (post-pre difference), $\rho=0.40$, $p<0.05$.

5.6 Discussion

Knowledge results indicated that participants knew only half the answers before using the SG, and that playing the SG over a 1-week period significantly increased their knowledge about the considered safety topics. Analysis

of each of the possible answers to the questions showed an improvement in all but two of the answers addressed by the game. The greatest improvement was found for the answer concerning smoke: before playing the SG only 5% of participants mentioned smoke among the dangers in aircraft evacuations, while after one week the percentage of participants who mentioned it was 68%. There was only one answer (about unruly passengers) for which the increase was not statistically significant, and another one (about panic) where we surprisingly found a statistically significant decrease. Probably most participants associated panic to aircraft emergencies before using the game, but one week later some of them focused more on the physical dangers they saw in the game, and panic was not portrayed in any way by the game. There was an improvement on the question that was not directly addressed by the game (bending down in smoke), but it was not significant, probably because the game did not allow participants to give a bend command to the character, and thus practice the correct behavior. In general, we observed that the answers with higher improvements tended to be those related to the behaviors that participants could practice more in the SG.

Since we did not ask participants to play as much as possible, but we instead asked them to play as much as they wanted in a naturalistic setting, the results concerning maximum level reached and playing time were encouraging in terms of the capability of the SG to keep participants engaged. The presence of several levels at increasing difficulty had likely contributed to increase playing time and maximum level reached. Indeed, all participants succeeded in the first easy levels and most of them were engaged by the increasing challenge, so they kept refining their skills to complete also the more complex levels. As a result, they spent time on the SG and performed (respectively avoided) the correct (respectively wrong) behaviors in the game several times to succeed. Since the answers with higher improvements were those related to behaviors that were practiced more, these results suggest that the presence of several levels at increasing difficulty could play a role in improving knowledge about correct behaviors, which in turn is fundamental to support attitude change towards them.

The SG received positive ratings for all except one of the constructs from the EGameFlow questionnaire. Participants agreed that the goals were clear and that the game provided prompt feedback (average ratings to goal clarity and feedback constructs were above 6 in a 1 to 7 scale). Both these results are important in the context of attitude change because they suggest that the game clearly conveyed the message concerning what to do in an aircraft emergency, and that it promptly provided feedback about correct and wrong behaviors performed by participants in the game. Notably, most of the statements that contributed to the feedback construct were about the immediateness of feedback. Therefore, the fast pace of the game, due to the time limit and the presence of danger and unruly passengers, could have highly contributed to the results. Participants also agreed that they remained concentrated on the game, and that it helped them to improve their knowledge (average ratings to concentration and knowledge constructs were between 5 and 6). One could have hypothesized that the use of humorous, fun, cartoon game elements from the arcade genre could have been distracting and thus could have prevented participants to elaborate the serious message of the game, but these results instead suggest that participants remained concentrated and perceived the game useful for the serious purpose of improving knowledge. In addition, participants agreed that the level of challenge and its increase were appropriate (average rating above 5). This suggests that our SG effectively employed the typical arcade game element of having levels at increasing difficulty. Finally, participants agreed that they felt a sense of control over the game, but in this case the average rating was closer to the neutral value (between 4 and 5). The only construct whose average rating was lower than the neutral value was immersion, but this could be due to the fact that the game was played on small displays. This explanation would be supported by studies that showed how the type of display used can affect perceived immersion, for a concise survey see (Buttussi and Chittaro, 2018).

The analysis of correlations of EGameFlow constructs showed that engagement played a role on both the knowledge improvement construct and the specific item about participants' motivation to learn more. In particular, it highlighted how both of them correlated with three different EGameFlow constructs, namely feedback, autonomy, and immersion. First, the more positively participants rated game feedback, the higher they rated knowledge improvement and the motivation item. These results suggest that the participants who better perceived the outcomes of their correct and wrong behaviors in the game, also felt more that they improved their knowledge, and were more motivated to further learn about the safety topic. Higher ratings for knowledge improvement and motivation were found also for participants who felt more sense of control over the game. A possible explanation for these correlations could be that participants who had more difficulties to understand why they failed some levels, and thus perceived less sense of control, were possibly frustrated and were less motivated to continue learning. However, further studies are needed to specifically address this correlation and support this explanation. Finally, participants who felt more immersed in the game, gave higher ratings to knowledge improvement and motivation. These results are particularly interesting because immersion ratings tended to be low, so there is much room for improvement. For example, experimenting with arcade game elements in virtual reality could be promising.

The correlations between motivation to learn more and participant's knowledge measured in pre-test and post-test were also interesting: the less one knew about aviation safety before playing, or the more knowledge one gained by playing the game, the higher the rating of the motivation to learn item. The first correlation suggests that the SG was useful to motivate those who needed it most, i.e. those who probably would not have followed the correct behaviors in an aircraft emergency because they did not know about them. The second correlation adds that those who learned most from the game were also more motivated to keep learning about the safety topic, indicating a positive attitude.

6. Conclusions and future work

The use of arcade games elements for attitude change in gamification and SGs has been scarcely studied. This paper contributed to advance knowledge about the topic, by studying the immediate and 1-week effects of a SG that uses such elements to engage players, increase their knowledge, and foster attitude change in aviation safety.

Overall, we showed that the SG could be more engaging than the traditional approach (safety card), and more effective to increase self-efficacy and perception of vulnerability to the considered risk. The game engaged participants also when used over a 1-week period in naturalistic settings, and was effective to increase knowledge about correct behaviors. Finally, we showed a role of engagement on attitudes, including agreement on correct behaviors and motivation to learn more.

The discussions in the paper provided an outline of the role that arcade game elements could have played on the results:

- The presence of several game levels of increasing difficulty offered an appropriate challenge that contributed to engagement, as confirmed by the time spent and the maximum level reached in the naturalistic settings, and by the positive ratings that participants gave to the EGameFlow challenge construct.
- The fast pace, due to the levels time limit and the presence of danger and unruly passengers, allowed the SG to provide feedback that was perceived as immediate and was appreciated by participants as

confirmed by the ratings they gave to the feedback construct of EGameFlow. Feedback, in turn, positively correlated with knowledge improvement and motivation to learn more.

- The use of humorous, fun, cartoon characters did not distract participants, as confirmed by the high ratings they gave to the EGameFlow concentration construct. However, the humorous approach might have been the cause of the lack of significant changes in severity.

The post-pre difference in severity was statistically significant in a study that evaluated a SG that used fear appeals for emergency preparedness in the domain of terror attack preparedness (Chittaro and Sioni, 2015). There was an increase in severity also in a SG that used fear appeals in the aviation safety domain, but the difference between the game and a safety card for this measure was only close to significance (Chittaro, 2016). Overall, these studies, together with (Chittaro and Buttussi, 2015; Buttussi and Chittaro, 2018), showed that the approach based on fear appeals can have a positive effect on engagement and on different measures related to attitude change (e.g., self-efficacy, locus of control, knowledge). However, fear appeals may also cause counterproductive defensive reactions (Ruiter, 2001) and not be appropriate for some users. For example, while some passengers who are comfortable with flying may underestimate the risks concerning aircraft emergencies and benefit from an approach based on fear appeals, people suffering from fear of flight may be frightened by SGs using fear appeals and also erroneously reinforce the wrong belief that aircraft emergencies are frequent. Similarly, while people can enjoy movies or games that depict aircraft disasters when they are at home, such materials are usually not included in the in-flight programs of airlines. When a fear-based approach may not be appropriate, the lighter, humorous approach based on arcade game elements that we proposed in this paper may be a valid alternative, since we found that it can engage users and have a positive effect on attitude change (including self-efficacy and knowledge about correct behaviors) despite the use of fun, cartoon characters instead of realistic, fearful simulations. Additional studies considering specific users and context (e.g., categorizing passengers based on the level of fear of flight) are required to assess when each approach is more appropriate.

More studies are also needed to extend findings to other safety topics and to other domains. For example, we are working on a new SG about procedures to open emergency exits that requires a very fast response to presented situations (i.e. players have to apply the appropriate procedure to open different types of emergency exits within a few seconds for each). This new SG will allow us to further test some of the arcade game elements employed in the present SG (e.g. the fast pace) as well as additional game elements (e.g. an endless runner approach with a continuous sequence of exits that is interrupted on the first error). Additional studies will be needed to assess the effects of arcade game elements in different domains.

Finally, since immersion promisingly correlated with motivation, and was one of the aspects of engagement in which the game could be improved more, we plan to explore the effects of arcade game elements also using immersive virtual reality, although this requires a thorough adaptation of the arcade game genre to take into account the different interaction techniques employed in immersive virtual reality.

Acknowledgments

This research was partially supported by a grant of the Federal Aviation Administration (FAA).

Nicola Zangrando (HCI Lab, University of Udine) developed the game used in the two studies.

References

Backlund, P., Engström, H., Hammar, C., Johannesson, M., Lebram, M., 2007. Sidh - a Game Based Firefighter Training Simulation, in: Proceeding of 11th International Conference on Information Visualization. IEEE, pp. 899–907. doi:10.1109/IV.2007.100

Bandura, A., 1997. Self-efficacy: the exercise of control. Freeman, New York, NY, USA.

Bandura, A., 2001. Social cognitive theory: an agentic perspective. Annu. Rev. Psychol. 52, 1–26. doi:10.1146/annurev.psych.52.1.1

Baranowski, T., Buday, R., Thompson, D.I., Baranowski, J., 2008. Playing for Real. Video Games and Stories for Health-Related Behavior Change. Am. J. Prev. Med. 34. doi:10.1016/j.amepre.2007.09.027

Bogost, I., 2011. Persuasive Games: Exploitationware. Blog post on Gamasutra [WWW Document]. URL https://www.gamasutra.com/view/feature/134735/persuasive_games_exploitationware.php (accessed 9.26.17).

Burrows, C.N., Blanton, H., 2016. Real-World Persuasion From Virtual-World Campaigns. Communic. Res. 43, 542–570. doi:10.1177/0093650215619215

Buttussi, F., Chittaro, L., Ranon, R., Verona, A., 2007. Adaptation of Graphics and Gameplay in Fitness Games by Exploiting Motion and Physiological Sensors, in: SG 2007: Proceedings of the 8th International Symposium on Smart Graphics, LNCS. Springer Berlin / Heidelberg, pp. 85–96.

Buttussi, F., Chittaro, L., 2018. Effects of different types of virtual reality display on presence and learning in a safety training scenario. IEEE Trans. Vis. Comput. Graph. 24, 1063–1076. doi:10.1109/TVCG.2017.2653117

Cafazzo, J.A., Casselman, M., Hamming, N., Katzman, D.K., Palmert, M.R., 2012. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J. Med. Internet Res. 14, e70. http://dx.doi.org/10.2196/jmir.2058

Chittaro, L., 2016. Designing Serious Games for Safety Education: “Learn to Brace” versus Traditional Pictorials for Aircraft Passengers. IEEE Trans. Vis. Comput. Graph. 22, 1527–1539. doi:10.1109/TVCG.2015.2443787

Chittaro, L., Buttussi, F., 2015. Assessing Knowledge Retention of an Immersive Serious Game vs. a Traditional Education Method in Aviation Safety. IEEE Trans. Vis. Comput. Graph. 21, 529–538. doi:10.1109/TVCG.2015.2391853

Chittaro, L., Sioni, R., 2012. Turning the classic snake mobile game into a location-based exergame that encourages walking, in: Bang, M., Ragnemalm, E.L. (Eds.), Persuasive 2012, LNCS 7284. Springer-Verlag, Berlin Heidelberg, pp. 43–54. doi:10.1007/978-3-642-31037-9_4

Chittaro, L., Sioni, R., 2015. Serious games for emergency preparedness: Evaluation of an interactive vs. a non-interactive simulation of a terror attack. Comput. Human Behav. 50, 508–519.

Cohen, B.H., 2001. Explaining Psychological Statistics, 2nd ed. Wiley, Hoboken, NJ, USA.

Consolvo, S., Klasnja, P., McDonald, D.W., Avrahami, D., Froehlich, J., LeGrand, L., Libby, R., Mosher, K., Landay, J.A., 2008. Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays, in: UbiComp 2008: Proceedings of the 10th International Conference on Ubiquitous Computing. ACM Press, New York, NY, USA, pp. 54–63. doi:http://doi.acm.org/10.1145/1409635.1409644

Connolly, T.M., Boyle, E. a., MacArthur, E., Hainey, T., Boyle, J.M., 2012. A systematic literature review of empirical evidence on computer games and serious games. Comput. Educ. 59, 661–686. doi:10.1016/j.compedu.2012.03.004

Corbett, C.L., McLean, G.A., Cosper, D.K., 2008. Effective Presentation Media for Passenger Safety I: Comprehension of Briefing Card Pictorials and Pictograms. Final Report DOT/FAA/AM-08/20. Washington, DC, USA.

Csikszentmihalyi, M., 1990. Flow: The Psychology of Optimal Experience. Harper & Row, New York, NY, USA.

De Hoog, N., Stroebe, W., de Wit, J.B.F., 2008. The processing of fear-arousing communications: How biased processing leads to persuasion. Soc. Influ. 3, 84–113. https://doi.org/10.1080/15534510802185836

DeSmet, A., Van Ryckeghem, D., Compernolle, S., Baranowski, T., Thompson, D., Crombez, G., Poels, K., Van Lippevelde, W., Bastiaensens, S., Van Cleemput, K., Vandebosch, H., De Bourdeaudhuij, I., 2014. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev. Med. (Baltim). 69, 95–107. doi:10.1016/j.ypmed.2014.08.026

Deterding, S., Dixon, D., Khaled, R., Nacke, L., 2011. From Game Design Elements to Gamefulness: Defining "Gamification," in: Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11. ACM Press, New York, NY, USA, pp. 9–15. doi:10.1145/2181037.2181040

Deterding, S., 2012. Gamification: Designing for Motivation. Interactions 19, 14–17. https://doi.org/10.1145/2212877.2212883

Fu, F.-L., Su, R.-C., Yu, S.-C., 2009. EGameFlow: A scale to measure learners' enjoyment of e-learning games. Comput. Educ. 52, 101–112. doi:10.1016/j.compedu.2008.07.004

Fuchslocher, A., Niesenhaus, J., Krämer, N., 2011. Serious games for health: An empirical study of the game "Balance" for teenagers with diabetes mellitus. Entertain. Comput. 2, 97–101. doi:10.1016/j.entcom.2010.12.001

Gnauk, B., Dannecker, L., Hahmann, M., 2012. Leveraging gamification in demand dispatch systems. In: Proceedings of the 2012 Joint EDBT/ICDT Workshops. ACM, New York, NY, USA, pp. 103–110.

Hamari, J., Koivisto, J., Sarsa, H., 2014. Does gamification work? - A literature review of empirical studies on gamification. Proc. Annu. Hawaii Int. Conf. Syst. Sci. 3025–3034. doi:10.1109/HICSS.2014.377

Hamari, J., 2017. Do badges increase user activity? A field experiment on the effects of gamification. Comput. Human Behav. 71, 469–478. doi:10.1016/j.chb.2015.03.036

Laamarti, F., Eid, M., El Saddik, A., 2014. An overview of serious games. Int. J. Comput. Games Technol. doi:10.1155/2014/358152

Lin, J.J., Mamykina, L., Lindtner, S., Delajoux, G., Strub, H.B., 2006. Fish'n'Steps: Encouraging Physical Activity with an Interactive Computer Game, in: UbiComp 2006: Ubiquitous Computing, Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 261–278. doi:10.1007/11853565_16

Liu, Y., Alexandrova, T., Nakajima, T., 2011. Gamifying intelligent environments. In: Proceedings of the 2011 International ACMWorkshop on Ubiquitous Meta User Interfaces. ACM, New York, NY, USA, pp. 7–12.

Massung, E., Coyle, D., Cater, K., Jay, M., Preist, C., 2013. Using crowdsourcing to support pro-environmental community activism. In: Proceedings of the 2013 ACMSIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 371–380.

Mau, G., Silberer, G., Constien, C., 2008. Communicating brands playfully. Int. J. Advert. 27, 827–851. doi:10.2501/S0265048708080293

Mayer, I., Wolff, A., Wenzler, I., 2013. Learning efficacy of the 'hazard recognition' serious game: A quasi-experimental study, In: Proceeding of 4th International Conference on Serious Game Development and Applications, LNCS 8101, Springer: Berlin, pp. 118–129.

Muir, H., Thomas, L., 2004. Passenger education: past and future, in: Proceedings of the 4th Triennial International Aircraft Fire and Cabin Safety Research Conference.

National Transportation Safety Board, 2000. Emergency Evacuation of Commercial Airplanes. Safety Study NTSB/SS-00/01. Washington, DC, USA.

Nicholson, S., 2012. A User-Centered Theoretical Framework for Meaningful Gamification, in: Proceedings of Games+Learning+Society 8.0 (GLS 8.0). ETC Press, pp. 223–230.

Petty, R.E., Wegener, D.T., 1998. Attitude change: Multiple roles for persuasion variables, in: Gilbert, D.T., Fiske, S.T., Lindzey, G. (Eds.), The Handbook of Social Psychology. McGraw-Hill, pp. 329–390.

Rapp, A., 2017. Designing interactive systems through a game lens: An ethnographic approach. Comput. Human Behav. 71, 455–468. doi:10.1016/j.chb.2015.02.048

Robertson, M., 2010. Can't play, won't play. Blog post on Hide and Seek [WWW Document]. URL http://hideandseek.net/2010/10/06/cant-play-wont-play/ (accessed 9.26.17).

Rogers, R.W., 1983. Cognitive and physiological processes in fear appeals and attitude change: A revised theory of Protection Motivation, in: Cacioppo, J.T., Petty, R.E. (Eds.), Social Psychophysiology: A Sourcebook. Guilford Press, New York, NY, USA, pp. 153–176.

Rose, K.J., Koenig, M., Wiesbauer, F., 2013. Evaluating success for behavioral change in diabetes via mHealth and gamification: MySugr's keys to retention and patient engagement. Diabetes Technol. Ther. 15, A114. http://dx.doi.org/ 10.1089/dia.2012.1221

Ruiter, R.A.C., Abraham, C., Kok, G., 2001. Scary warnings and rational precautions: A review of the psychology of fear appeals. Psychol. Health 16, 613–630. doi:10.1080/08870440108405863

Sakamoto, M., Nakajima, T., Alexandrova, T., 2012. Value-based design for gamifying daily activities, in: Proceedings of the 11th International Conference on Entertainment Computing (ICEC'12). Springer-Verlag, Berlin, Heidelberg, pp. 421–424.

Seaborn, K., Fels, D.I., 2015. Gamification in theory and action: A survey. Int. J. Hum. Comput. Stud. 74, 14–31. doi:10.1016/j.ijhcs.2014.09.006

Schwarzer, R., Jerusalem, M., 1995. Generalized Self-Efficacy scale, in: Weinman, J., Wright, S., Johnston, M. (Eds.), Measures in Health Psychology: A User's Portfolio. Causal and Control Beliefs. Nfer-Nelson, Windsor, UK, pp. 35–37.

Stinson, J.N., Jibb, L.A., Nguyen, C., Nathan, P.C., Maloney, A.M., Dupuis, L.L., Gerstle, J.T., Alman, B., Hopyan, S., Strahlendorf, C., Portwine, C., Johnston, D.L., Orr, M., 2013. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J. Med. Internet Res. 15, e51.

Sweetser, P., Wyeth, P., 2005. GameFlow: A Model for Evaluating Player Enjoyment in Games. ACM Comput. Entertain. 3, 3A. doi:10.1145/1077246.1077253

Wise, K., Bolls, P.D., Kim, H., Venkataraman, A., Meyer, R., 2008. Enjoyment of Advergames and Brand Attitudes. J. Interact. Advert. 9, 27–36. doi:10.1080/15252019.2008.10722145

Yang, M., Roskos-Ewoldsen, D.R., Dinu, L., Arpan, L.M., 2006. The Effectiveness of “in-Game” Advertising: Comparing College Students’ Explicit and Implicit Memory for Brand Names. J. Advert. 35, 143–152. doi:10.2753/JOA0091-3367350410

Zyda, M., 2005. From visual simulation to virtual reality to games. IEEE Comput. 38, 25–32. doi:10.1109/MC.2005.297