NOTICE: This is a paper that was accepted for publication in the IEEE Transactions on Visualization and Computer Graphics. It is not the final published version. The DOI of the definitive version available in IEEE Explore is http://dx.doi.org/10.1109/TVCG.2015.2443787
Designing Serious Games for Safety Education: “Learn to Brace” vs. Traditional Pictorials for Aircraft Passengers

Luca Chittaro

Abstract—Serious games for safety education (SGSE) are a novel tool for preparing people to prevent and/or handle risky situations. Although several SGSE have been developed, design and evaluation methods for SGSE need to be better grounded in and guided by safety-relevant psychological theories. In particular, this paper focuses on threat appeals and the assessment of variables, such as safety locus of control, that influence human behavior in real risky situations. It illustrates how we took into account such models in the design and evaluation of “Learn to Brace”, a first-of-its-kind serious game that deals with a major problem in aviation safety, i.e. the scarce effectiveness of the safety cards used by airlines. The study considered a sample of 48 users: half of them received instructions about the brace position through the serious game, the other half through a traditional safety card pictorial. Results showed that the serious game was much more effective than the traditional instructions both in terms of learning and of changing safety-relevant perceptions, especially safety locus of control and recommendation perception.

Index Terms—serious games, safety education, aviation, mobile devices, threat appeals.

1 INTRODUCTION

Current safety education is mainly based on traditional media such as printed materials, videos, and oral briefings. Unfortunately, the effectiveness of these methods of instruction is often very limited in safety education. For example, research in the aviation domain shows that passengers’ attention to traditional safety instructions is poor at best and even passengers who do pay attention have little knowledge and understanding of the information received [18][44][55]. Reports of the US Federal Aviation Administration (FAA) [18] as well as other sources in the literature [6][44][55] thus recommend to design new tools that could both: (i) educate passengers about aviation safety in more creative ways [19][44], and (ii) be available outside the aircraft, increasing passengers’ exposure to safety education materials, e.g. safety education exhibits at all airports [6] or interactive digital disks that could be passed out at airports, air shows, and public events [19].

Serious games, i.e. video games to further training and education objectives [63], are considered as a novel tool for safety education in different domains, e.g., road safety [40], fire safety [49], work safety [27], home safety [37], military safety [52]. They can provide players with interactive, realistic virtual experiences that could be more engaging and easier to comprehend than traditional media. Moreover, they could be played on the user’s own digital devices. Therefore, they could meet the two above mentioned recommendations. Unfortunately, the positive effects of serious games for safety education (SGSE, for short) are not clear yet, because they are typically evaluated with simplistic methods or not evaluated at all (see Section 2).

This paper aims at advancing knowledge about SGSE in different directions. First, we propose to ground SGSE design on psychological theories of how people respond to information and recommendations about risks. Second, we follow a more thorough SGSE evaluation methodology that includes: (i) comparison with traditional education materials used in the considered safety domain, (ii) extension of game effects assessment to psychological variables that influence human behavior in real risky situations. Third, we illustrate how we applied the proposed ideas to the design and evaluation of a first-of-its-kind serious game in aviation safety education. Our study found that the game was much more effective than a traditional aviation safety education method, both in terms of learning and of changing safety-relevant perceptions. The ideas for SGSE design and evaluation proposed in this paper are not limited to the aviation domain, and lend themselves to applications in other safety domains.

The paper is organized as follows. In Section 2, we discuss previous work on SGSE and their evaluation as well as research in the specific domain of aviation safety education, highlighting how our work differs from it. Section 3 motivates and gives a theoretical grounding to the additional variables we propose to consider in the design and evaluation of SGSE. In Section 4, we illustrate in detail the serious game created by applying the proposed ideas. Section 5 and 6 respectively
illustrate the user evaluation and its results, while Section 7 discusses the results and concludes the paper.

2 RELATED WORK AND MOTIVATIONS

2.1 Serious Games for Safety Education (SGSE) and their Evaluation

An increasing number of serious games aims at preparing people to prevent or handle risky situations. The topics in such SGSE span several domains, e.g. submarine safety [52], on-board ship firefighting [58], neutralization of saboteur-created dangers on military ships [54], terrorist threat prevention in ports [45], procedures for inmates' safe transfer in prisons [20], safe operation of combustibles and home heating equipment [37], prevention of work accidents [27][41][57] and home accidents [57], evacuations of buildings [49][51], road safety [40], fire safety [12][43][46], and street safety [15].

Unfortunately, the considerable effort spent on developing SGSE is not accompanied by a comparable effort in investigating if they positively affect their players. For most SGSE, no user evaluation is provided and the literature. The available SGSE evaluations, e.g. [15][40][41][43][46][52], are instead limited for the following two main reasons.

First, they do not compare SGSE with the materials traditionally employed to deliver the same safety instructions, and they are thus unable to say if SGSE can actually improve current levels of safety education. Evaluation of serious games must instead include comparison with traditional instruction methods before making claims about effectiveness of the approach [25].

Second, they concentrate on measuring knowledge acquisition in SGSE players, which is just one of the necessary outcomes that enable an individual to actually perform the right actions in a real risky situation. Connoly et al. [16] and Graesser et al. [26] discuss how playing a serious game can have perceptual, cognitive, behavioral, and affective impacts, e.g. knowledge acquisition as well as motivational outcomes. The need for a thorough and rigorous evaluation of the different outcomes is stressed also by a recent review of serious game assessment [4]. As clearly pointed out by major theories of human behavior - e.g., Planned Behavior [1], Social Cognitive [3], and Protection Motivation [47] theories - different people with similar knowledge and skills may perform differently depending on variables such as motivation, perceived efficacy of the received recommendations or perceived personal control on the situation. The design and evaluation of SGSE has to consider such aspects.

For these reasons, the evaluation methodology we use in this paper focuses on both (i) including a comparison with traditional safety instructions, and (ii) assessing variables that can influence safety performance.

SGSE could also benefit from an extension of the psychological models they employ to influence users into learning safety actions. Current SGSE typically exploit operant conditioning [23], i.e. user’s responses that the game wants to encourage are reinforced through positive feedback (e.g., the ability to proceed further in the game, receiving points and badges, being praised with positive comments,…), while those that need to be discouraged are possibly punished with aversive feedback (getting stuck at a given point in the game or taken back to the beginning of a level, losing points and lives, receiving criticism for the errors made,…). When properly implemented, such approach can include desirable features such as enabling players to explore hypothetical situations, instantly observe the link between cause and effect [23][26], providing immediate feedback, and possibly showing the consequences of the chosen behavior. Immediate feedback can also facilitate engagement and self-efficacy [26]. Moreover, a game can simulate consequences in vivid ways that contribute to make them memorable [13].

However, game design that is merely guided by an operant conditioning strategy does not provide a sufficient guarantee of success because it disregards the mechanisms and variables that lead people to accept safety recommendations and change their attitudes concerning the considered risky situations. The methodology used in this paper will thus extend SGSE design based on operant conditioning with additional theories, described in Section 3.

2.2 Improving Safety Education with Serious Games: the Case of Aviation Safety

To discuss the limitations of traditional safety instructions, which are common to many domains, we use aviation safety education as a paradigmatic case study, for three main reasons. First, compared to other safety education domains that would require lengthy introductions, it is a convenient case study for the majority of readers, who are likely to be aircraft passengers and already have first-hand experience of how airlines deliver safety instructions. Second, aviation safety education is an important issue that concerns billions of passengers in the world (airlines flew 3.1 billion passengers in 2013 [2]). Third, studies of the effectiveness of traditional aviation safety education are available in the literature as summarized in this section.

The purpose of aviation safety education is to provide aircraft passengers with accurate cabin safety knowledge and cultivate positive passengers' attitudes to affect appropriately their behavior when an emergency occurs. As shown by [7], the level of passengers’ aviation safety education does affect their knowledge, attitudes and behavior. Safety education can lead passengers to efficient behavior and being responsible for their own safety. Therefore, improving their safety education increases probability of survival in emergencies [44][55].

Current approaches to aviation safety education rely on the safety card in the seat pocket and the pre-flight briefing to which passengers are exposed on the aircraft.
Unfortunately, as stressed by FAA reports [18] as well as other authors in the literature [6][44][55], passenger attention to safety cards and briefings is poor at best, and even the few passengers who do pay attention have little knowledge and understanding of the information received. As a result, lack of passenger preparedness is a major cause of death and injury that could be preventable in aircraft accidents. For these reasons, the above cited sources recommend to design new tools that could both: (i) educate passengers about aviation safety in more creative ways [19][44], and (ii) be available outside the aircraft, increasing passengers’ exposure to safety education materials, e.g. creating safety education exhibits at all airports [6] or interactive digital disks that could be passed out at airports, air shows, and public events [19].

SGSE are a natural candidate for trying to meet both recommendations. First, they can provide players with interactive, realistic experiences that could be more engaging and easier to comprehend than traditional materials, and include individual feedback about user’s errors and how to correct them. Second, they can be used at home (or anywhere, if available on mobile devices), whenever and how many times players want.

However, to the best of our knowledge, the only previous attempt at building a serious game for aviation safety education was the one we described in [9]. In that work, we created a simple 3D world that allowed users to partially try a specific pre-scripted aircraft evacuation scenario. Progress in the scenario was determined at some checkpoints in which the player had to choose an action, e.g., retrieving his/her luggage or leaving it on the plane. Choosing the right action made players progress in the scenario, otherwise they were left at the checkpoint and the question was asked again. The negative consequences of wrong choices were described only by a very brief text and not simulated with graphics and sound, because we did not want to scare the player. The user interface for choosing actions was only a text menu, and the player had to press keys on the keyboard to choose from the menu. The study on 26 users showed that playing the scenario (which required about 3 minutes to complete) improved to some extent knowledge of the evacuation procedure and self-efficacy (the participants’ belief in their ability to carry out a behavior [3], evacuation in this case). The experience had an effect on users’ perception of emergency landing risk: while vulnerability did not change, there was an undesired significant decrease in severity. This was possibly due to the fact that we specifically designed the experience not to be scary and we did not show any kind of adverse effects on the players’ avatar.

The game we present in this paper differs from the one in [9] in several ways. First, it focuses on educating passengers about how to accurately perform a specific action (assuming a brace position), while the previous prototype focused on an abstract procedure in which complex actions (such as assuming the brace position) reduced to simply selecting a single menu item. Second, it aims at creating more playful dynamics, in which users do not simply choose items from a text menu, but can actively play with their avatar body, posing it in a wide range of different postures and seeing how this affects it in an accident. Third, it aims at providing much richer feedback: (i) while the previous prototype did not show any negative consequences, we now fully simulate them with realistic graphics and sound, and (ii) after showing the scary consequences, we provide hints to help the player avert them. Fourth, since this time we aimed at making the game available for public campaigns, we devoted particular care to obtaining a graphic and audio quality higher than that of typical research prototypes. The methodology proposed in this paper led us to change game dynamics with respect to our previous work as well as extend the study of game effects to important constructs from safety-relevant psychological theories, introduced by the next section.

3 Extending the Theoretical Basis of SGSE Design and Evaluation

SGSE need to warn people about threats and foster awareness about how to prevent them and/or act properly when they occur. Unlike other educational games and materials, they must focus on risky situations that can have negative consequences on the health of the player and/or other human beings, including the possibility of serious injury and death. In addressing these peculiar aspects of SGSE, we explore the adoption of the following constructs (summarized in Figure 1) and theoretical models that are important in safety but currently unused in SGSE design and evaluation.

3.1 Threat Appeals

In psychology, a communication strategy that concentrates on warning people about risks and the negative consequences of specific actions with the aim of changing the recipient’s attitudes and/or behavior is traditionally called threat appeal [59][60][48], and more recently threat appeal, because fear is one possible reaction in response to a threatening stimulus [39]. To understand better when and how threat appeals work or fail, the literature proposes theories of how individuals respond to information and recommendations about risk, and are motivated to protect themselves from risks. Witte's Extended Parallel Process Model [59] and Roger’s Protection Motivation Theory [47] are two leading theoretical models. Both point out that recipient’s perceptions of the presented risk as well as recommendation are key factors for the effectiveness of a threat appeal. The role of specific variables in determining such perceptions, extensively studied in psychological research and confirmed by independent meta-analyses [24][60], can be summarized as follows.

First, the message must threaten the individual by highlighting the severity of the risk and the vulnerability
of the individual to it. If the individual perceives the risk as not severe or perceives himself/herself as not vulnerable to it, the models predict that (s)he will not feel threatened and will not be motivated to consider how to cope with the risk.

Second, the message must highlight that there is an effective action that can avert the risk and that the individual is capable of that action. If the individual perceives the action as ineffective or perceives himself/herself as not capable of doing it, the models predict that (s)he will try to reduce negative emotions induced by threat through processes such as risk denial and defensive reactions, which are detrimental to learning proper safe behavior for that risk.

Meta-analyses have shown that high-threat messages are in general more effective than low-threat ones in changing recipient’s attitudes [24][48][60]. However, since they did not differentiate among different manipulations of severity and vulnerability in the threat messages, de Hoog et al. [21] conducted a meta-analysis of possible differential effects, pointing out that severity is more important than vulnerability in attitude change.

Bringing threat appeal models into SGSE design, perceived risk severity can be affected by how the game portrays the negative consequences of the risk on the player and the threatening audio-visual stimuli used to that purpose. Vulnerability can instead be highlighted by situating the virtual risk experience in contexts (e.g., an accurate reproduction of the cabin of an airliner in the case of aviation safety) that users can recognize as those in which they can find themselves in the real world.

Following threat appeal models, SGSE have to show what would happen to the individual if the risk is not avoided, employing threatening stimuli to highlight negative consequences (e.g., the game proposed in this paper shows and emphasizes bodily harm on the player’s avatar) as well as the effectiveness of the recommended action (e.g., posing the avatar properly on the seat will leave it unharmed, while errors in the position will produce different kinds of bodily harm to the avatar). The perception of being capable to perform the proper actions (e.g., putting the head on the seat in front) should not be impaired by complex game interfaces that could make users perceive real-world actions as more complex than they are. For example, in our game, the interface should allow users to put easily and naturally the avatar head on the seat. A simple interface should also support a seamless gameplay that is not interrupted by interaction complexities, to keep players engaged and focused on the content.

3.2 Safety Locus of Control

In addition to threat appeals, we extend our attention to a construct, safety locus of control, which is an important indicator of the possibility that users will follow the learned recommendation when actually faced with the risk in the real world. In general, locus of control (LOC) can be defined as the degree to which a person perceives that the outcomes of the situations (s)he experiences are under his/her personal control [35]. Given a specific situation, an individual’s LOC can have an internal orientation (the individual perceives that she can exert control over the outcome of the situation) or an external orientation (the individual perceives that the outcome of the situation is due to external factors, such as fate, chance or the actions of other persons).

Several studies highlighted the importance of LOC in risky situations, showing that internal LOC is a predictor of safer attitudes and behaviors. For example, in road safety, Hoyt [33] showed that car passengers with an internal orientation are more likely to wear seat belts, while Montag and Comrey [42] related drivers’ internal LOC with safer driving. Interestingly, a recent study [34] showed that drivers’ LOC can be influenced by training and by observer feedback, and the change in drivers’ LOC can predict change in driving behavior. This suggests that the simulations one can explore in SGSE and the feedback they provide to players are worth studying also as techniques to change safety LOC.

Wuebker [61] focused on LOC in the industrial safety domain, proposing the Safety Locus of Control Scale. Results of her study indicated that externally oriented employees had significantly more accidents than employees with an internal safety LOC orientation. Moreover, accidents and injuries suffered by externally oriented employees were more serious than those of internally oriented ones. Jones and Wuebker [38] studied hospital workers, confirming the relation of the Safety Locus of Control Scale with occupational accidents in that domain too. Hunter [35] adapted the Safety Locus of Control Scale to measure aviation safety LOC in pilots and found that civil aviation pilots with a more internal orientation were involved in fewer hazardous events. A study of airline pilots by You et al. [62] reinforced these findings, showing that safety LOC influenced safe operation behavior. Hunter and Stewart [36] extended
the investigation to U.S. Army aviators, developing the Army Locus of Control Scale and finding significant associations that were consistent with previous research. In particular, aviators with high internal LOC experienced fewer accidents than aviators who were low on that construct.

Improving safety LOC appears to be particularly important also in the domain of passenger safety. Indeed, passengers tend to look at aviation emergencies with attitudes that are more consistent with an external rather than an internal orientation, e.g. shifting the responsibility of their safety to the cabin crew [44] or falsely believing that most aircraft accidents are unsurvivable [53]. This way of thinking is dangerous because external orientation has negative effects on safety attitudes and behavior, and is unfounded for different reasons. First, the crew cannot provide individual assistance to every passenger in an emergency, due to workload and time constraints. Moreover, crew members could be injured or incapacitated, and this would require passengers to take an even more active role to survive. Second, passengers’ pessimistic beliefs about survivability are contradicted by facts: a survey of commercial jet airplanes accidents conducted by Boeing [5] indicated that the majority of accidents is survivable, and a recently released FAA report [8] confirmed and reinforced that conclusion.

For the reasons summarized in this section, a design goal of SGSE in any domain should be to produce greater improvements in safety LOC than traditional safety education materials. To the best of our knowledge, this paper is the first to investigate if SGSE can be effective in achieving this goal.

4 THE SERIOUS GAME

A fundamental action that passengers can take to contribute to their survival in aircraft accidents is to assume an appropriate “Brace for Impact” position that can significantly reduce injuries sustained [53]. The purpose of our serious game is to allow players to become familiar with all the details of assuming a brace position as well as to change their attitudes, in particular their safety LOC, concerning emergency landings. The game, called “Learn to Brace”, is meant for use in on-line public campaigns, so that users can conveniently play it on their digital devices to familiarize with the position well before boarding a plane.

In designing “Learn to Brace”, we organized gameplay into the following four steps (interested readers can also freely try the game first hand on different platforms [28][29][30][31]). First, players pose their avatar, a 3D virtual passenger. They see it seated in the cabin of a flying aircraft, from a third-person perspective (Figure 2). We carefully reproduced the cabin of a typical airliner (including environmental sounds) to create a situation that any user who has taken a flight can immediately recognize as personally familiar, possibly contributing to perception of vulnerability. Moreover, the game interface allows users to pose quickly and easily the avatar, possibly affecting perception of recommendation simplicity. The interface consists of four distinct icons (Figure 2), each one
associated to a different movable part: head, hands, seat belt, and feet. The player can freely drag each icon to move easily the associated part. For example, if the user drags the “hands icon” towards the top/bottom or the left/right of the screen, (s)he will drag the hands of the passenger respectively up/down and forward/backward in the cabin environment. The arms of the passenger will follow the hands in a physically consistent way, as if the player was a “puppeteer” who controls the avatar by pulling strings. We adopted such “puppeteer” metaphor for the interface because it should be immediately familiar to users. Moreover, drag-and-drop interaction (through which users individually drag each of the 4 icons) is typical of many computer applications and should thus require no learning effort. The game invites players to pose the passenger in a position they believe to be safe for an emergency landing, but players are completely free to pose the character in dangerous positions.

Second, when the player clicks the “Crash” button in Figure 2, the game simulates a hard emergency landing in real-time. The simulation is not pre-scripted and is physically based so each crash can be different from a previously seen one, depending on the initial body position. The aim of the simulation is to vividly show the consequences of assuming a wrong position on the passenger body (possibly affecting risk severity perception) as well as allow the player to learn experientially which positions do not produce such negative consequences (possibly affecting perception of recommendation efficacy). Figure 3 shows three instants of a specific simulation.

Third, the game displays a slow-motion replay, designed to look similar to videos of car crash tests (possibly enhancing perception of risk severity and recommendation efficacy, as for the real-time simulation). Slow-motion allows players to appreciate details that cannot be noticed in the real-time simulation. Indeed, real-time simulation occurs in a very short time as an impact in real life, making it impossible to perceive all the details of the impact dynamics on the moving avatar body. To make slow-motion replay more dramatic, we show it in greyscale and we highlight in red the parts of the body damaged by the impact (Figure 4). Moreover, ominous sounds (crash landing and breaking bones) are reproduced at the proper instants.

Fourth, the illustration of the outcome of each simulation is completed by providing players with a detailed damage report and individual feedback that includes explicit recommendations (Figure 5). We highlighted in red the externally visible damage to the virtual body (right side of Figure 5). Then, inspired by the recently proposed idea of using medical imaging visualizations of internal damage to body parts for health persuasion [50], we enriched the report with x-ray visualizations (left side of Figure 5) that show internal damage to the affected body parts. As discussed in Section 3, we followed threat appeal models that highlight how scaring the user can be a good tactic as long as the game also presents a simple and effective way of averting the depicted negative consequences. To do so, for each negative consequence on the passenger’s body, the game provides a recommendation in the form of a short and clear hint about a simple action (possibly affecting recommendation simplicity) that the user can easily carry out in the real world to avoid that consequence (left side of Figure 5). This is also an implicit incentive to retry again the game to see if one can do better next time. A prominent “Retry” button (Figure 5, bottom right) allows players to conveniently go back to the avatar-posing phase and make changes to its position. This way, the simulation enables users to see for themselves how they can survive the impact by following the recommendations about simple actions that are under their control (position of head, hands, seat belt, and feet). This can possibly improve their safety LOC as well as perception of recommendation efficacy.

To choose a correct brace position for the game, we
first analyzed official advice about brace positions [56], which derived from dynamic impact tests conducted in the 80’s and 90’s. However, the FAA has recently conducted new impact tests to review effectiveness of brace positions and determine if the ones in use are still appropriate for today’s passenger seats [53]. The need for new tests was prompted by the injuries sustained by passengers in several recent commercial airline accidents (e.g., US Airways flight 1549) [53]. In “Learn to Brace”, we thus adopted the optimized brace position recommended by the new impact studies [53].

To implement “Learn to Brace”, we used C# and the Unity 3D game engine.

5 USER STUDY

To analyze the effectiveness of the “Learn to Brace” game, we carried out a between-groups study that compared it with the traditional instruction method used in aviation safety cards, i.e. the pictorial, and assessed all the variables introduced in Section 3.

The pictorial we used (shown in Figure 6) followed graphic framing choices (subject and surroundings, field size, camera angle) taken from brace pictorials displayed in official civil aviation recommendations ([56], Appendix 4). It was also consistent with the graphic choices of safety cards employed by different airlines, which provide a color representation of the cabin environment surrounding the passenger. We also made sure to use the same graphics of the serious game, to avoid introducing confounding differences in the way the cabin, the passenger, and the brace position were depicted in the safety card and in the game. The pictorial was printed on glossy paper and had a 5x8 cm size, a typical pictorial size in airline safety cards.

5.1 Participants and Design

We recruited 48 participants (23 male, 25 female) through personal contact. Participants were volunteer university students and people from various occupations who received no compensation. Their age ranged from 19 to 55 (M=29.88, SD=12.49).

Video game use was assessed by asking participants to rate frequency of use 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). Ten users reported that they never played video games, 14 played once a month or less, 11 played several times a month, 8 played several times a week, and 5 played every day.

Frequency of air travel was assessed by asking participants...
participants for their number of flights in the last two years, as in [18]. Each flight had to be counted individually, e.g. a round trip from airport A to airport C via a connection through airport B resulted in 4 flights. Answers ranged from 0 to 15 (M=2.88, SD=3.32).

The study followed a between-groups design, with type of safety education material (Safety Card or Serious Game) as independent variable. The 48 participants were assigned to the two groups in such a way that: (i) the proportion of men and women was very similar (12M and 12F in the Safety Card group, 11M and 13F in the Serious Game group), (ii) the two groups were very similar in terms of age, video game use, and frequency of air travel. Lack of significant differences between the two groups was confirmed by independent samples t-test for age (t(46)=.52, p=.61), and by Mann-Whitney test for video game use (U=287.5, Z=-.01, p=.99) because the variable was ordinal.

5.2 Measures

To evaluate factorial validity of questionnaire items used to measure the constructs described in Section 3, we performed exploratory factor analyses with principal component extraction and Varimax rotation. The following sections include results of each analysis.

5.2.1 Knowledge

To measure participants’ knowledge about the brace position, we used four questions that asked them to describe where and how to position (i) hands, (ii) feet, (iii) safety belt, and (iv) head. To avoid suggesting possible answers (e.g., as in a multiple-choice questionnaire), we asked participants to answer the questions orally and we recorded the answers. Knowledge was measured as the number of correctly answered questions and thus ranged between 0 and 4.

Participant’s knowledge measured before exposure to the considered safety education materials was low and very similar in the two groups: mean value was 1.37 (SD=.65) in the Safety Card group and 1.10 (SD=.58) in the Serious Game group. This very small difference in initial knowledge between the two groups was not statistically significant (t(46)=1.64, p=.11).

5.2.2 Safety Locus of Control

To measure participants’ safety LOC concerning emergency landing situations, we adapted items (6 for internal LOC, 6 for external LOC) from the Aviation Safety Locus of Control Scale [35] by changing the context from the pilot’s to the passenger’s one. Factor analysis revealed that one of the items for external LOC (“In an emergency landing, injuries and deaths are caused by unsafe equipment and poor safety regulations”) did not load clearly on neither internal nor external LOC. This was likely due to the fact that it was the only item that referred to a legal topic (“poor safety regulations”) that is familiar to aviation professionals but not to passengers. This probably made it difficult for participants to interpret correctly the item. After removal of the unclear item, factor analysis was carried out on the 11 items listed in Table 1, for each of the two datasets (questionnaires filled respectively before and after exposure to the education materials). Post-hoc indicators of data and sampling adequacy for the factor analysis (Kaiser-Meyer-Olkin, KMO, and Bartlett’s Test of Sphericity) showed that KMO was greater than .60 (.66 in the first dataset, .85 in the second), and Bartlett’s test was significant (p<.001 in both datasets). The analysis confirmed the intended two-factor structure that explained respectively 49.6% of variance in the first dataset and 67.5% of variance in the second dataset.

Answers to the 6 internal LOC (resp. the 5 external LOC) items were averaged to form a reliable scale; Cronbach’s alpha for internal (resp. external) LOC was .73 (resp .76) before exposure to the education materials, and .85 (resp .90) after exposure.

Safety LOC before exposure to the education materials was very similar: internal LOC was 4.41 (SD=1.05) in the Safety Card group and 4.30 (SD=.80) in the Serious Game group. This initial very small difference was not statistically significant (t(46)=.44, p=.66). External LOC was respectively 3.89 (SD=1.13) and 3.80 (SD=1.18). This very small difference was not statistically significant (t(46)=.28, p=.79).

5.2.3 Risk Perception

We measured risk perception by using the 6 items employed by [22], rated on a 7-point scale (1=not at all, 7=very), changing the name of the considered risk into “emergency landing”. Participants rated risk severity on three items which respectively asked how severe, harmful, and serious the consequences of an emergency

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SAFETY LOCUS OF CONTROL ITEMS AND SCORING KEY</strong></td>
</tr>
<tr>
<td>-------------------------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Key</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>I</td>
</tr>
</tbody>
</table>

<sup>a</sup>I = Internal; E = External.
landing would be. Cronbach’s alpha was .93 before exposure to the education materials, and .94 after exposure. The three items for vulnerability asked instead how vulnerable respondents perceived themselves to be with respect to an emergency landing; how high they thought their risk of being involved in an emergency landing was; and how high the probability of suffering personal negative consequences from an emergency landing was. Cronbach’s alpha was .65 before exposure to the education materials, and .66 after exposure.

Factor analysis of the two datasets (questionnaires filled respectively before and after exposure to the education materials) showed that KMO was greater than .60 (it was .78 in both datasets), and Bartlett’s test was significant (p<.001 in both datasets). The analysis confirmed the intended two-factor structure that explained respectively 75.9% of variance in the first dataset and 75.1% of variance in the second dataset.

Risk perception before exposure to the education materials was very similar: vulnerability was respectively 4.04 (SD=1.31) in the Safety Card and 4.00 (SD=1.39) in the Serious Game group; severity was respectively 5.40 (SD=1.58) and 4.92 (SD=1.59). These initial small differences between the two groups are not statistically significant (vulnerability: t(46)=.18, p=.86; severity: t(46)=1.06, p=.29).

5.2.4 Recommendation Perception

We measured perceived efficacy and simplicity of the recommendations by having participants rate their level of agreement with 6 items on a 7-point Likert scale (1=not at all, 7=very). The items made the following statements about the recommendations provided by the education materials: “They are useful for my safety”, “They will allow me to face effectively an emergency landing”, “By following them, I can greatly reduce my probability of getting hurt in an emergency landing”, “They are simple to learn”, “They are easy to remember”, “They are easy to carry out”. Answers to the first (resp. last) three items were averaged to form a reliable scale (Cronbach’s alpha=.92, resp. .85) for perceived efficacy (resp. perceived simplicity).

Factorial analysis showed that KMO was greater than .60 (it was .71), Bartlett’s test was significant (p<.001), and confirmed the intended two-factor structure that explained 83.3% of variance in the dataset.

5.3 Hypotheses

Following the considerations discussed in Sections 3 and 4, we formulated the following hypotheses: (i) player’s knowledge about the brace position should improve more with the serious game than the safety card, (ii) the serious game should bring a larger improvement in safety LOC (i.e. a larger increase in participants’ internal LOC and a larger decrease in external LOC) than the safety card, (iii) perceived efficacy of the materials should be consistent with the hypothesized LOC improvement, with participants perceiving the safety recommendations as more effective when received through the serious game than the card. For risk perception, we did not expect differences in vulnerability between educational materials, because such perception is more related to the traveling habits of participants. Moreover, the graphical representation of the aircraft environment and passenger was the same in the two conditions. On the contrary, the vivid and possibly scary visualization of bodily harm to the player’s avatar, which is central to “Learn to Brace”, made us hypothesize a larger increase in severity in the Serious Game group than the Safety Card group. We expected “Learn to Brace” to show an opposite trend with respect to our previous simple game [9] that did not show bodily harm and lead to a significant decrease in severity (see Section 2), and we were interested in exploring to what extent this might occur.

5.4 Procedure

Participants in the Serious Game (resp. Safety Card) group were told they were going to try a video game (resp. a safety card) meant to illustrate the position passengers should assume during an emergency landing. They were told they could use the game (resp. safety card) for as much time as they wanted, with no minimum or maximum limits. First, participants filled the demographic, risk perception and safety LOC questionnaires, and answered the knowledge questions. Then, participants in the Safety Card group examined the safety card. Participants in the Serious Game group played “Learn to Brace” on a 15.6 inches LCD display with in-built stereo speakers, without receiving any training or illustration of the game from the experimenter. This was done to check that the user interface was intuitive and immediately usable as planned. All instructions for playing the game were in a brief text displayed as a starting screen that said: “You are going to face an emergency landing! Click and drag the 4 yellow icons with the mouse to choose a position you think is safe. When you’re done, press CRASH!”. The model of virtual passenger used in the study was the one depicted in Figures 3, 4, and 6. When participants decided to stop playing (resp. examining the safety card), they filled the recommendation perception, risk perception, safety LOC questionnaires, and answered the knowledge questions.

6 RESULTS

Variables described in Sections 5.2.1, 5.2.2, and 5.2.3 were measured before and after exposure to the aviation safety education materials. Following Cohen’s recommendations about the most powerful analysis method for before-after measurements [14], we analyzed each of those variables with a between-groups ANCOVA in which the before value was introduced as a covariate, the after value as dependent variable, and the type of education material was the independent variable. Since variables described in Section 5.2.4 were instead measured only after exposure to the education materials, we performed a one-way ANOVA analysis. The results for vulnerability showed no statistically significant differences between the two groups (vulnerability: t(46)=.18, p=.86; severity: t(46)=1.06, p=.29).
materials, we used between-subjects ANOVA for them.

6.2.1 Knowledge

Means for knowledge scores are shown in Figure 7. The effect was significant ($F(1,45)=12.57, p<.001$) and its size was large ($\eta^2_{p}=.41$). In the Serious Game group, mean knowledge score was 3.92 (SD=.28), a value extremely close to the best possible one, and 22 out of 24 group members obtained maximum knowledge score (4), while the remaining two members obtained 3. On the contrary, mean knowledge score in the Safety Card group was 3.00 (SD=.93), only 8 out of 22 members obtained maximum score, and scores for the remaining 14 members included also values as low as 1 and 2.

6.2.2 Safety Locus of Control

Figure 8 shows the means for safety LOC. The effect was significant, its size was medium to large for internal LOC ($F(1,45)=9.45, p=.004$, $\eta^2_{p}=.17$) and medium for external LOC ($F(1,45)=7.09, p=.011$, $\eta^2_{p}=.14$). The Serious Game group improved its internal LOC, reaching a mean value of 4.91 (SD=.77), while there was no improvement in the Safety Card group after exposure (M=4.32, SD=1.33). Similarly, the Serious Game group improved its external LOC, which decreased to a mean value of 3.42 (SD=.98), while there was no improvement with Safety Card (M=4.15, SD=1.44).

6.2.3 Risk Perception

Figure 9 shows the means for vulnerability and severity. As expected, changes in vulnerability were very small, and there was no significant difference between the groups ($F(1,45)=1.61, p=.21$, $\eta^2_{p}=.04$). For severity, the mean increased in the Serious Game group while it remained unchanged in the Safety Card group, and the effect was close to significance ($F(1,45)=3.02, p=.08$, $\eta^2_{p}=.06$). Since these results did not reach significance, we carried out a post-hoc power analysis for a 0.05 alpha, using the observed sample effect size as the basis of the population effect: statistical power was low (.24 for vulnerability, .40 for severity).

6.2.4 Recommendation Perception

Figure 10 shows the means for simplicity and efficacy of the safety recommendations. Perceived simplicity was high in both groups, and the slight difference in favor of Serious Game was only close to statistical significance ($F(1,46)=3.07, p=.08$, $\eta^2_{p}=.06$) with a low (.40) statistical power (computed as in 6.2.3).

The difference in perceived efficacy was instead large, with a mean rating equal to 6.15 (SD=7.92) for Serious Game and 4.61 (SD=1.69) for Safety Card, the effect was significant ($F(1,46)=16.46, p<.001$), and the effect size was large ($\eta^2_{p}=.26$).
6.2.5 Considerations about Exposure Time

We analyzed how much time participants devoted to examine the education materials in the two groups. The difference was statistically significant (F(1,46)=76.35, p<.001), and effect size was large (η²=.26): mean time was 275.63 s (SD=136.66) in the Safety Card group and 29.00 s (SD=21.08) in the Serious Game group. Most participants (22) in the Serious Game group played for a time that varied between 145 s and 389 s (M=241.38, SD=71.84), while the remaining two played for a considerably larger time (579 s and 728 s). Similarly, the number of tries of the simulation was for most players (22) between 2 and 5 (M=3.18, SD=.85), while for the remaining two players it was respectively 6 and 8.

Considering the above time difference, and the fact that length of exposure to education materials could affect learning, we re-run the ANCOVA for knowledge, adding exposure time as a covariate. The analysis confirmed the significance of the difference between Safety Card and Serious Game (F(1,144)=11.00, p=.02) and a medium to large effect size (η²=.20), controlling for exposure time.

Finally, it is interesting to note that, in the Serious Game group, exposure time was positively correlated (r(24)=.57, p<.01) with participants’ age: the older the participants, the more time it took them to play the game. There was instead no correlation between age and exposure time in the Safety Card group.

7 DISCUSSION AND CONCLUSIONS

This paper pursued four main goals: (i) extending the theoretical grounding for the design and evaluation of SGSE, (ii) applying the proposed approach to the design of the first serious game for instructing about brace positions, (iii) following an evaluation method that, in addition to learning, considered important psychological constructs for the domain of safety, and (iv) thoroughly evaluating the game not just in isolation, but in comparison with traditional methods (safety card) used to instruct passengers about the brace position.

The results of the study show that a serious game designed as proposed can be a much more effective educational tool than traditional safety cards. Moreover, learning obtained with the serious game was extremely close to the best possible outcome: playing the game for an average time of just 4 minutes resulted in full learning of the safety knowledge for all but two players. Playing the game also improved important attitudes about emergency landings, increasing players’ internal safety LOC and decreasing external safety LOC. In other words, participants who played the serious game felt the outcomes of an emergency landing were more under their personal control after playing the game. This desirable outcome was instead not obtained in participants who used the safety card, which produced no safety LOC improvements at all.

The results also suggest that the game might be able to enhance the perception of risk severity, while the safety card produced no change in such perception. This result should be examined in contrast to our previously studied simple game [9] that caused a significant decrease in severity (as seen in Section 2). This different outcome obtained by the two games is consistent with the design choice of vividly showing the scary consequences of the accident on the passengers’ body in “Learn to Brace”, while deliberately omitting to show them in the previous game.

In addition to showing such negative consequences, the design proposed in this paper aimed at presenting recommendations in a way that could make players perceive them as simple and effective, as prescribed by threat appeal models. Results show that safety recommendations were perceived to be simple in both groups, but there was instead a large difference in the perception of their efficacy: playing the serious game resulted in a much larger perception of recommendation efficacy than using the safety card.

Overall, the obtained positive results seem to support the game design approach we have used in this paper. We showed players possibly scary simulations of negative consequences of wrong actions, but we were also careful to always accompany them with simple hints to avert the depicted consequences. We provided players with a user interface that all participants were able to use immediately and allowed them to try the recommendations and check their effectiveness with a clear simulation.

A limitation of the study lies in the fact that all participants were occasional flyers (average of 2.88 flights in the last two years). Future studies should extend the sample to frequent flyers. Research on comprehension of airline safety cards carried out with occasional as well as frequent flyers [18] showed that for some (but not all) topics illustrated by pictorials, occasional as well as frequent flyers [18] showed that for some (but not all) topics illustrated by pictorials, the familiarity of high-flight-time passengers with safety briefings and cards allowed them to obtain better comprehension results. The same paper remarks that such finding confirms that aviation safety education materials need to be designed for novice passengers.

Another limitation of our sample is that most participants were familiar with video games and were relatively young adults (average age was 29.88). It would thus be interesting to repeat the study with people who are completely unfamiliar with video games and/or belong to other age groups (older adults, children) to test if the obtained results still hold for these different samples.

Finally, while the study focused on attitude change and knowledge gain, it did not consider behavior change. To explore this aspect, one might involve participants in real-world simulations of emergencies in an aircraft cabin (such facilities are available in some aerospace research centers). Nevertheless, even in this case, users would always be aware that they are not actually involved in a real accident, an issue that is
common to other safety training research. Moreover, it must be noted that such real-world simulations of aircraft emergencies present ethical issues, due to a significant potential for injuries, including serious ones, to the involved participants [17].

After the user study, we made two changes to the game: (i) a few cosmetic improvements to the virtual human in the game, which resulted in the two virtual humans shown in Figures 2 and 5, and (ii) the addition of a second game level that deals with the less common, special case in which a passenger is seated too far from the surface in front of him/her and cannot thus place the head on it (see Figure 11). Then, in April 2014, we publicly released “Learn to Brace”. The game is available on mobile and desktop platforms: for desktop computers (PCs and Macs), we released it as a Facebook app [30]; for mobile devices (smartphones and tablets), we released it for Android [28], Apple iOS [29], and Windows Mobile [31], on the respective stores. From April 2014 to April 2015, “Learn to Brace” was played by more than 22’000 users (65.2% iOS, 26.5% Android, 7.8% Windows Mobile, 0.5% Facebook). Public accessibility of the game will allow us to conduct remote evaluations, involving large, international, and varied samples of players, and to study how they use the game in more naturalistic ways than laboratory studies.

Developing the game for multiple platforms was an important step to achieve the goal of creating safety education tools that passengers can use outside the aircraft (see Sections 1 and 2.2). In addition, the game obtained better results than traditional education materials, meeting the other recommended goal. However, grounding SGSE design on threat appeals also presents a limitation. The possible arousal of fear due to the vividly simulated emergency makes the game content not suitable for some passengers on an aircraft. This issue generally affects any content provided by in-flight entertainment systems: for example, while people can watch and enjoy movies that depict aircraft disasters, this is generally inappropiate during certain phases of flights (e.g., take off and landing). A preliminary, publicly accessible example of this second line of research is provided by our recent LifeVest prototype [32], in which we are focusing on teaching passengers how to wear a life preserver. In LifeVest, the interaction is very similar to “Learn to Brace”: the user moves the hands of the virtual passenger to make it wear the life preserver correctly. However, we appeal to different emotions, e.g., we are trying to introduce humor (the avatar can make funny remarks or movements in response to player’s right or wrong choices), and frame the game as a time challenge in which the player has to carry out all the right actions in as less time as possible (which is also a correct message to send, because passengers must be able to wear life preservers in a short amount of time).

In both lines of research, we will progressively develop a game version of each safety topic whose inclusion is mandatory in airline safety cards [56]. The game versions of the different safety instructions will eventually be integrated into a single, larger game. For the games based on threat appeals, the integrated game will simulate entire aircraft accidents, following the timeline provided by official reports on real representative accidents, and will be inspired to action/survival game genres. Such game could be suitable to train people at home in evaluating the possible different situations that occur in real aircraft accidents and come up with action plans to survive the different scenarios.

The non-threatening games will be instead integrated into an interactive, game-based version of a full safety card, in which each topic is illustrated by a game. This interactive application could be used on-board on the passenger’s own digital devices (the latest FAA policy is to extend personal electronic devices use by passengers to the various phases of flights) or in-flight entertainment systems.

It must be noted that (i) the two different types of games are not mutually exclusive, but can be used in synergy (one on-board and one on the ground), appealing to different emotions and exploiting different
game genres to make safety knowledge memorable, (ii) on-board games for aviation safety education will likely co-exist with traditional aviation safety briefings and cards, because not every passenger is necessarily attracted by video games, and first-time passengers who have never played the game before boarding cannot be expected to do it very quickly before taking off.

Finally, we are broadening our research focus to experiment the design and evaluation approach of this paper in other safety domains. In particular, we are focusing on evacuation preparedness for mass emergencies in public places.

ACKNOWLEDGMENTS

This research is supported in part by a grant from the US Federal Aviation Administration (FAA). We are grateful to Cynthia Corbett and Mac McLean for their feedback and encouragement.

Nicola Zangrando (HCI Lab, University of Udine) carried out 3D modeling and computer programming activities for the development of the game.

REFERENCES

[30] HCI Lab, University of Udine. Learn to Brace, Facebook App for PC and Mac, [Online]. Available:


