Pre-test influences on the effectiveness of digital-game based learning: A case study of a fire safety game

Anissa All a,⁎, Barbara Plovie b, Elena Patricia Nuñez Castellan a,c, Jan Van Looy a

a Department of Communication Sciences, iMinds–MICr-Ghent University, Belgium
b HOWEST University of Applied Sciences, Belgium
c Department of Data-analysis, Ghent University, Belgium

ARTICLE INFO

Article history:
Received 9 March 2016
Received in revised form 23 December 2016
Accepted 28 May 2017
Available online 31 May 2017

Keywords:
Digital game-based learning
Effectiveness assessment
Solomon four-group design
Practice effect
Pre-test sensitization

ABSTRACT

In recent years, critiques have been formulated regarding current evaluation methods of digital game-based learning (DGBL) effectiveness, raising doubt with regard to the validity of certain results. A major issue of contention is whether or not a pre-test should be administered, gauging for baseline measures of knowledge that are targeted using an educational intervention.

The present study aims to explore the advantages and disadvantages of adding a pre-test in DGBL effectiveness research. For this purpose, an effectiveness study of a fire safety training in a hospital was conducted using a Solomon four-group design. The experimental groups received a game-based intervention (N = 65) of which one group received a pre- and a post-test (n = 34) and one group only received a post-test (n = 31). The control groups received traditional classroom instruction (n = 68), of which one group received a pre- and a post-test (n = 39) and one group only received a post-test (n = 29). No main effect of testing was found. However, an interaction effect between pre-test and intervention was detected. Subjects who have received a pre-test in this group score significantly higher (p < 0.05) on the post-test than subjects in the traditional classroom group who did not receive a pre-test. This was not the case in the game group.

When the administration of a pre-test influences the control group’s receptivity to the intervention, but not that of the experimental group, results of an effectiveness study may be biased. Hence comparison of post-test scores of different treatments in pre-test/post-test designs may be problematic. This is an important finding in the context of DGBL effectiveness research as the presence of a pre-test may artificially inflate the learning outcomes of the control condition. Therefore, further research should take this into account and look for possible solutions to solve this discrepancy. However, in the present study, we were able to show that the game was highly effective, as both game groups still outperformed the slide-based group that received a pre-test. The Solomon four group design has thus shown its added value and more effectiveness studies on DGBL implementing this design are required in order to further validate these results.

© 2017 Published by Elsevier Ltd.

⁎ Corresponding author. Korte Meer 7-9-11, 9000 Gent, Belgium.
E-mail address: anissa.all@ugent.be (A. All).

http://dx.doi.org/10.1016/j.compedu.2017.05.018
0360-1315/© 2017 Published by Elsevier Ltd.
1. Introduction

The interest in using digital games as instructional tools has increased strongly over the past decade. Digital game-based learning (DGBL) refers to the usage of the entertaining power of digital games to serve an educational purpose (Prensky, 2001). The goal of DGBL is thus twofold: it has to be fun/entertaining and it has to be educational (Bellotti, Kapralos, Lee, Moreno-Ger, & Berta, 2013). There are several reasons why scholars believe that digital games are considered an appropriate medium for instruction. Firstly, digital games possess certain attributes that can positively influence the learner’s motivation to start and persist in the educational intervention. Secondly, games also contain certain attributes that allow the implementation of certain learning paradigms. DGBL can be motivating in two ways. Firstly, DGBL can be implemented to ‘seduce’ the learner by gameplay to allocate his/her attention to the learning content (Ritterfeld, Weber, Fernandes, & Vorderer, 2004). Interactivity is one of the main characteristics of game-based learning resulting in higher attention during the activity and consequently, deeper processing of the content (Ritterfeld et al., 2004). Secondly, DGBL can stimulate intrinsic motivation to engage in the training due to the enjoyable experience it provides (Garris, Ahlers, & Driskell, 2002). This means, for instance, that learners wish to finish the game training because it is fun or because they wish to achieve in-game goals rather than because they are obliged to finish the training. Intrinsic motivation is, in turn, related to higher levels of engagement, performance, higher quality of learning and lower levels of dropout (Ryan & Deci, 2000).

The added value of DGBL is, however, not only related to its motivational power, but its learning mechanisms also fit well within modern theories of effective learning proposed by educationalists and psychologists (Boyle, Connolly, & Hainey, 2011). Digital games allow for the implementation of constructivist theories of learning (Boyle et al., 2011; Rooney, 2012). Constructivism relies on the assumption that learning is a process in which learners’ knowledge and skills are constructed by making sense of their experiences. In constructivist learning theory, the learner is an active learner as opposed to a passive one receiving and processing information provided by an instructor (Hein, 1991). Main constructivist learning mechanisms that underpin the instructional potential of DGBL are situated learning, experiential learning and problem-based learning (Boyle et al., 2011; Rooney, 2012). Games can enable situated learning, according to which learning is context-dependent and needs to occur in the context of the authentic learning environment to which the learning applies (environment, actions, situations and actors) (Ladley, 2010). An authentic learning environment is one that replicates what the learner would experience in a real-world situation. Learning is thus a result of the interaction of mental processes with the physical and social environment (Clancey, 1991). In certain cases such as emergency situations, a simulation of that authentic environment is the best alternative solution for providing this situated learning experience (Ladley, 2010). Digital games have the ability to provide this authentic environment, both regarding the simulation of the actual physical environment, events and consequences of actions made in this simulated world.

Digital games also enable an experiential learning experience, according to which experiences are a source of learning and one learns by doing (Kolb, 1984). According to Kolb, an experiential learning experience is a cyclical process which consists of four phases. The first phase is the concrete experience, followed by the second phase, reflective observations, where the learner observes and reflects on this experience. Based on these observations and reflections, the learner draws conclusions and makes hypotheses and generalizations on how this acquired knowledge can be used in other situations, which is called abstract conceptualization. The final phase in this cyclical process is active experimentation, where the learner tests these hypotheses by experimenting and applying the acquired knowledge. This process also occurs while playing video games, requiring “…a constant cycle of hypothesis formulation, testing, and revision. This process happens rapidly while the game is played, with immediate feedback” (Van Eck, 2006, p. 5).

Digital games also offer the potential to provide a problem-based learning experience (Van Eck, 2015, pp. 13–28), where a particular problem is presented to the learners and knowledge and skills are acquired during the process of solving this problem (Savery & Duffy, 1995). Problem solving is a mechanism that often occurs in digital games, by means of goals or missions a player has to accomplish (Kiili, 2005).

1.1. Empirical evidence on DGBL effectiveness

DGBL has been implemented in various sectors such as defense, education, corporate training, health and wellbeing, and communication (Backlund & Hendrix, 2013). Concomitantly there has been a growing interest and production in research into DGBL’s effectiveness (Mayer et al., 2014; Wouters, Van Nimwegen, Van Oostendorp, & Van Der Spek, 2013). In recent years, a significant amount of research assessing the effectiveness of DGBL has been published (Hainey et al., 2014; Hwang & Wu, 2012). Results regarding its effectiveness are, however, mixed. While some meta-analyses have found that DGBL is more effective than non-game instructional methods regarding learning gains, others have found non-significant differences (Backlund & Hendrix, 2013; Clark, Tanner-Smith, & Killingsworth, 2014). The same inconsistency is found regarding motivational outcomes (Clark et al., 2014; Wouters et al., 2013).

These mixed results are at least in part due to the variety of shapes that DGBL presents in terms of topics and game genres (Kirriemuir & McFarlane, 2004). Another important factor has been the heterogeneity in study designs for assessing their effectiveness however (Authors). Research designs differ in several respects including use of a control group, activities presented in the control group(s), implementation of DGBL (stand-alone vs. in a broader program), outcome measures to assess effectiveness, statistical techniques to quantify learning outcomes and the administration of a pre-test (Girard, Ecalle, & Magnan, 2013; Authors). Moreover, methodological issues have been brought forward regarding published effectiveness
research on DGBL (Clark, 2007; Clark, Tannser-Smith & Killingsworth, 2014; Girard et al., 2013; Authors), which sometimes lacks rigorous assessment (Clark, 2007; Clark et al., 2014; Connolly, 2014). For instance, studies are frequently being implemented without a strict control of potential threats to their internal validity, such as the addition of training materials to the intervention (e.g., required reading, exercises) or the lack of a standardized protocol for instructors (e.g., procedural help, guidance only during the intervention). Moreover, authors regularly fail to mention whether or not self-developed tests have been piloted, which leads to uncertainty with regard to the reliability and validity of results (Brom, Sisler, Buchtovà, Klement, & Levčík, 2012, pp. 41–53). Another important methodological issue is that it is difficult to replicate published DGBL effectiveness studies given that authors often do not provide sufficient information on how the intervention — both in the experimental and control condition — has been implemented (Sitzmann, 2011; Authors). Detailed information on procedure is indispensable, however, in order to gain insight in whether the gains that are reported are a consequence of the different methods and not due to other circumstantial factors that differed between conditions (Randel, Morris, Wetzel, & Whitehill, 1992) and to be able to replicate studies (Authors).

Considering these methodological limitations, a more systematic approach that can serve as a guideline for quality assessment is required for researchers willing to conduct effectiveness studies in this field (Mayer et al., 2014). For this purpose, research into preferred study designs is required. In the present study, we aim to investigate whether or not a pre-test of knowledge should be administered, as the absence of a pre-test is one of the main criticisms of DGBL effectiveness studies (Clark, 2007; O’Neil, Wainess, & Baker, 2005; Authors) and studies without a pre-test are consequently often omitted from meta-analyses on DGBL effectiveness (e.g., Clark et al., 2014; Girard et al., 2013).

1.2. Pre-test administration

Administration of a pre-test is a contentious topic as gauging for baseline measures of knowledge can provide additional data regarding participants but it can also threaten the validity of results. Adding a pre-test to the research design is useful as it allows researchers to control for pre-existing differences between the experimental and control group (Clark, 2007) and to compare progress (i.e., gain scores) as a result of the interventions (Gerber & Green, 2012). By adding pre-test scores to the analysis — for example when comparing gain scores or conducting repeated measures or analysis of covariance with pre-test scores as covariate — error variance is also reduced, resulting in a more precise estimate of the treatment effect, allowing for the use of statistically more powerful tests (Dimitrov, Philip, & Rumrill, 2003; Knapp & Schafer, 2009). Lastly, the addition of a pre-test allows the researcher to control for characteristics of drop-outs (Authors) so that potential biases with regard to representativeness of the sample can be reported. On the other hand, adding a pre-test can also ‘blur’ the real effect of the treatment. Firstly, administering a pre-test can result in ‘practice effects’, meaning that subjects who take the same test twice, may do better the second time, even if the intervention had not taken place (Crawford, Stewart, & Moore, 1989). In this case the effect is due to the pre-test, as it can offer participants additional exercise materials or item training (van Engenlburg, 1999). Hence progress due to the intervention and progress due to the practice effect cannot be isolated from each other. Moreover, pre-test sensitization can occur, referring to an interaction effect of the pre-test and the treatment (Braver & Braver, 1988; van Engenlburg, 1999). This means that subjects who have received a pre-test will be more sensitive to the intervention as compared to subjects who have not received a pre-test, resulting in higher scores on the post-test. For instance, when implementing the same test pre- and post-intervention during a short period of time, a pre-test can cue students on what should be remembered from the intervention (Randel et al., 1992). Consequently, one cannot know whether a positive effect as a result of the treatment would have been present if a pre-test had not been administered. In that case generalization of results from a pre-tested to an un-pretested sample is made impossible. This has resulted in researchers renouncing a pre-test when studying effectiveness of DGBL (e.g., Amory, 2010; Tsai, Yu, & Hsiao, 2012). To our knowledge, pre-test influences have hitherto never been studied in a DGBL context. One meta-analysis conducted by Wouters et al. (2013), however, has investigated the impact of experimental design (post-only vs pre-test-post-test design) on magnitude of effects found in DGBL effectives studies. No significant differences could be found between those designs regarding learning outcomes. Therefore, before making assumptions on the presence of a practice effect or pretest sensitization, this needs to be studied (Braver & Braver, 1988).

An experimental design that is proposed to investigate the issues of practice effects and pre-test sensitization is the Solomon four-group design (Solomon, 1949). A Solomon four group design is an experimental design with four conditions: two treatment and two control conditions. One treatment condition gets a pre-test before the intervention and in the other treatment condition, a pre-test is absent. The same applies to the control conditions. Table 1 provides a schematic overview of this design.

A meta-analysis on Solomon four group designs in the area of psychology conducted by Willson & Putnam, 1982 has shown that administration of a pre-test can result in elevated and biased post-test scores. This is especially the case for cognitive learning outcomes: 93% of the studies investigating pre-test effects in an intervention aimed at cognitive learning outcomes found higher scores among pretested participants compared to unpretested participants. For attitudinal outcomes, 62% of the studies investigating pre-test effects found higher scores among pretested participants compared to unpretested participants. Hence, they conclude that ‘there is a general pretest effect that cannot be safely ignored’ (p. 13) and further research regarding these effects is needed. However, in the last 2 decades almost no studies can be found related to this issue, especially not regarding educational interventions, let alone technology enhanced instruction. One study implementing a Solomon 4-group design in this field can be found (Arbaugh, 2000). In this study, an internet-based course was compared to a
Table 1
Schematic overview of the Solomon 4 group design.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Pre-test</th>
<th>Intervention</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment condition 1 (T1)</td>
<td>Yes</td>
<td>X</td>
<td>Yes</td>
</tr>
<tr>
<td>Control condition 1 (C1)</td>
<td>Yes</td>
<td>C</td>
<td>Yes</td>
</tr>
<tr>
<td>Treatment condition 2 (T2)</td>
<td>No</td>
<td>X</td>
<td>Yes</td>
</tr>
<tr>
<td>Control condition 2 (C2)</td>
<td>No</td>
<td>C</td>
<td>Yes</td>
</tr>
</tbody>
</table>

X stands for experimental treatment (subject of study, DGBL). O stands for observation. T stands for experimental/treatment group and C for control treatment (slide-based lecture).

traditional classroom course. Results brought forward by this study show that the post-test scores of the participants receiving a pre-test before the traditional class outperformed participants in the traditional class that did not receive a pre-test. In the internet-based course, no such differences could be found. This could imply a pre-test sensitization effect, but interactions between administration of a pre-test and treatment were not investigated.

The present study conducts a similar study as Arbaugh (2000) in the context of DGBL effectiveness research. More specifically, we aim at testing for a main effect of pre-test (i.e., pre-test effect) and an interaction effect between pre-test and treatment (i.e., pre-test sensitization) on learning outcomes.

2. Method

2.1. Design

A Solomon four-group design was implemented in order to assess the effectiveness of a digital game-based fire safety training among hospital personnel. Participants in the experimental condition received a digital game-based intervention and participants in the control group received the traditional slide-based lecture. Individual randomization of subjects was not possible in this study due to practical limitations, as staff needs to subscribe for the fire safety training (i.e., the hospital disposes of a large pool of staff such as nurses, cleaning personnel, doctors, etc. who work in shifts). Hence, randomization was implemented on a group level (i.e., a group was composed of people that subscribed for a safety training on the same date).

2.2. Stimulus material

2.2.1. Digital game-based fire safety training

The DGBL fire safety training was specially developed for the hospital whose personnel participated in the study. All hospital personnel (i.e., doctors, nurses, cleaning personnel, administrative staff, technical staff, etc.) is required to complete the fire safety training every year. Because the hospital has expanded over the years and is still expanding and personnel works in different shifts, it is becoming increasingly difficult to organize traditional training for everyone. Hence the decision to develop a digital game in cooperation with DAE research. The game consists of three mini-games or courses: ‘small fire’, ‘smoke’ and ‘blaze’. The minigames are 3 small interactive simulations where the player can earn coins for taking the correct steps (e.g., answering a multiple choice question on what is the correct action to take), performing the right action (e.g., alarm internally by calling the correct number) or performing the correct order of actions (e.g., steps to take to activate a fire extinguisher). During the minigame the players receive information regarding correct actions and procedures by means of information cards. These information cards can be consulted at every time during gameplay. When the player answers a question wrong, performs an action incorrectly or does not apply the right order of procedures, he/she can try again until he provides the correct answer. The more attempts the player needs to be correct, the less coins he/she earns. At the end of every minigame an overview of the earned coins is provided. After participants have completed these courses, they can also play a random ‘fire safety’ scenario, during which elements learned in the course can be practiced. In total, 6 different scenarios are available. A description of the game and gameplay can be found in Table 2. The game can be freely played on the following website: http://sgso.howest.be/het-serious-game/

2.2.2. PowerPoint

The PowerPoint lecture is instructed by the prevention manager of the hospital. This is the lecture that is currently being used as a fire safety training for the hospital personnel. This lecture applies the same structure as the game: small fire, smoke, gaze. Per type of emergency situation, the right steps and procedures are discussed. A fire extinguisher, a fire blanket ad a fire hose are the only extra materials used during the lesson, to show the staff how to use them (see ‘small fire’ in Table 2). This lecture was also used as a basis to define content treated in the game and contains exactly the same material as treated in the game. The slide-based lecture applies a passive learning approach.
### Table 2
Description of the fire safety training game.

<table>
<thead>
<tr>
<th>Minigame</th>
<th>Description</th>
<th>Learning elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small fire</td>
<td>A small fire has arised in a hospital room. The player has to take several steps when this happens (e.g., internal alarm and different options, extinguish fire and different options, etc.). Three types of learning elements are integrated in this minigame: answering a multiple choice question (e.g., what is the first step to take when there is a small fire?), perform an activity (e.g., alarming your colleagues) and applying the right order of actions (e.g., show the right order of actions to activate the fire extinguisher).</td>
<td>2 procedures for internal alarm (information is provided + two procedures need to be performed in game) 3 devices to use to extinguish fire Procedure to follow in order to extinguish fire with fire extinguisher (show the right order of actions) Procedure to follow in order to extinguish fire with fire blanket (show the right order of actions)</td>
</tr>
<tr>
<td>Smoke</td>
<td>Smoke is coming from a hospital room. The player has to take several steps when this happens (e.g., feel heat at the door, alarm internally, open the door, right position to open the door, what to do when fire does not extinguish etc.).</td>
<td>Alarm internally (activity needs to be performed by the player) Alarm externally (activity needs to be performed by the player and correct information needs to be provided) Right position to open the door (player needs to show the right position in the game) Extinguish fire (player needs to attempt to extinguish fire in the game) Procedure to evacuate mobile patients (player needs to execute evacuation)</td>
</tr>
<tr>
<td>Blaze</td>
<td>There is a blaze in a hospital room. The player has to take several steps when this happens (e.g., internal alarm, external alarm, evacuation in right order, etc.)</td>
<td>Alarm internally and externally (activities needs to be performed by the player and correct information needs to be provided) Procedure to evacuate wheelchair patient (player needs to show the correct steps to take) Procedure to evacuate immobile patient (player needs to show the correct steps to take)</td>
</tr>
<tr>
<td>Random scenario</td>
<td>This is a random scenario which consists of one of the events above.</td>
<td>See above.</td>
</tr>
</tbody>
</table>

### 2.3. Procedure

#### 2.3.1. Experimental groups

The experimental groups played the game in a conference room on one of the four campuses of the hospital during working hours. A maximum of six subjects could participate per session. When entering the conference room, subjects received an introduction by a researcher regarding the purpose of the study. Afterwards, the subjects either filled out the pre-test (experimental condition with pre-test) or started playing the game (experimental condition without pre-test).

The subjects played the game individually on a laptop computer with a headphone. During game play, two researchers were present providing procedural help, meaning that only technically oriented help was provided when there were issues with the computer or game play (i.e., no help regarding course materials). After the subjects completed all three courses and one scenario, a post-test was administered. In total, 18 game training sessions were organized; 9 included a pre-test and 9 did not.

#### 2.3.2. Control groups

The control groups equally received the slide-based lecture in a conference room on one of the four campuses. The slide-based lecture was given by either the prevention manager or another fixed employee from the prevention staff that was responsible for the fire safety training. The same procedures were followed regarding administration of pre-test and post-test as in the experimental groups. The subjects were instructed in groups of minimum 8 and maximum 20 people. In total, 6 slide-based lectures were organized, 3 included a pre-test and 3 did not. During every slide-based lecture, the same two researchers who were present during the DGBl intervention were present to check whether all topics discussed in the game were also discussed in the slide-based lecture using a topic list (see appendix A).

### 2.4. Participants

The present study was conducted in collaboration with the hospital AZ groeninge in Kortrijk (Belgium). In total, 152 subjects participated in the study. Eighty-three subjects participated in the experimental groups, of whom 42 received a pre-test and 41 did not receive a pre-test. Sixty-nine subjects participated in the control groups of whom 39 received a pre-test and 29 did not receive a pre-test. Nineteen subjects in the experimental group (8 who received a pre-test and 11 who did not receive a pre-test) were excluded from the analysis because log data showed that they either did not complete all three courses or they repeated a course several times. One participant in the control group was excluded from the analysis, because she did not speak Dutch and could not understand the questions in the test. In the end, 133 participants were retained for the analysis.
As can be seen in Table 3, randomization on a group level has led to a balanced group in terms of age and proportion of gamers, but not in terms of gender composition.

2.5. Measures

Three types of outcomes should be considered when assessing effectiveness of DGBL: learning outcomes, motivational outcomes and efficiency outcomes (Authors). DGBL is considered effective if it succeeds in achieving similar learning outcomes compared to more traditional methods, without significantly diminishing any of the others. In the present study, we have assessed performance as an indicator for cognitive learning outcomes, motivation towards the instructional material as an indicator for motivational outcomes and time exposed to intervention as an indicator for efficiency outcomes.

2.5.1. Cognitive learning outcomes

In order to assess performance, a test was developed by the researchers in cooperation with the prevention staff responsible for the fire safety training – the same staff who provided the slide-based lectures. The test had previously been implemented in a pilot study (N = 52) in an initial phase of development of the game. The test consisted of 18 open-ended questions, covering all topics that are treated in the interventions allowing for a maximum score of 40. The test assesses declarative and procedural knowledge. Examples of questions are: What is the first step you have to take when a small fire breaks out? How do you do this? What are the three steps you have to follow when evacuating patients? Which three steps do you have to take to evacuate a bedridden patient? Etc. The tests were corrected by two researchers. For this purpose, an evaluation form was developed in order to guarantee a standardized manner of correcting the tests. If there was uncertainty regarding the correctness of certain answers, the correctors discussed the response and agreed upon a score. The translated test and information regarding scoring can be found in appendix B.

2.5.2. Motivational outcomes

The Instructional Materials Motivation Survey (IMMS, Keller, 1987) was used to assess motivation towards the instruction method. We based ourselves on Huang, Huang, and Tschopp (2010) for the game version of the IMMS. The IMMS consists of 36 items, divided in 4 subscales: attention (i.e., gaining and keeping the learner’s attention), relevance (i.e., activities must relate to current situation or to them personally), confidence/challenge (i.e., activities cannot be perceived as too hard or too easy, which is also a prerequisite for an optimal game experience or game flow) and satisfaction/success (i.e., learners must attain some type of satisfaction or reward from the learning experience). The items were scored on a 5-point Likert scale, with 1 being ‘not true’ to 5 ‘very true’. The total score represents motivation towards the instructional material. The scores on the subscales give an indication as to the sub dimensions on which the intervention was either more or less successful (Keller, 2010).

A reliability analysis of our data showed an acceptable Cronbach’s alpha for subscales attention (α = 0.82) and satisfaction (α = 0.84), but not for subscales confidence (α = 0.50) and relevance (α = 0.67). Hence, we deleted confidence items 1 (‘When I first looked at the game/slides, I had the impression that it would be easy for me’) and 34 (‘I could not really understand quite a bit of the material in the game/slides.’) We have also deleted relevance item 26 (‘the game/lecture was not relevant to my needs because I already knew most of it’). With these items deleted, confidence (α = 68) and relevance (α = 0.70) have an acceptable Cronbach’s alpha.

2.5.3. Efficiency outcomes

Time management as an efficiency outcome refers to DGBL succeeding in reducing the timeframe needed to teach a certain content matter (authors). Hence, we have timed every separate slide-based lecture and retrieved individual information on total time spent on the DGBL intervention based on automated logging.

2.6. Data analysis

Assumptions for normality were checked by comparing the numerical value for skewness and kurtosis with the respective standard error (Field, 2009) and checking the Q-Q plot of standardized residuals of the dependent variable (Kutner, Nachtsheim, Neter, & Li, 2005). Since the post-test scores were not normally distributed (i.e., negatively skewed), a

<p>| Table 3 |</p>
<table>
<thead>
<tr>
<th>Control for balanced groups as a result of randomization on group level.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group with pre-test (n = 34)</td>
</tr>
<tr>
<td>Women</td>
</tr>
<tr>
<td>Age (mean)</td>
</tr>
<tr>
<td>Gamers</td>
</tr>
</tbody>
</table>
reversed square root transformation was applied on the post-test data for the analyses of variance. The difference scores were normally distributed, so no transformation was necessary for the paired samples t-test. To check for equality of variance, Levene’s test was used.

3. Results

Firstly, we will discuss the effectiveness of the DGBL intervention and secondly, we will discuss the influence of the pre-test on outcome results.

3.1. Effectiveness of the DGBL treatment

Two designs (pre-test post-test design and post-only) can be distinguished in our data (van Engelenburg, 1999) in order to assess the effectiveness of DGBL, we will conduct analyses on two datasets: one containing the participants receiving both a pre- and a post-test and one on the participants only receiving a post-test.

3.1.1. Pre-test post-test design

A paired samples t-test (N = 73) showed a difference between pre- and post-test scores for both the participants receiving a slide-based lecture, t(38) = 20.65, p < 0.01, r = 0.92, and the participants receiving the DGBL intervention t(33) = 14.57, p < 0.01, r = 0.87, showing that both produce a large learning effect. Table 4 provides an overview of the descriptive statistics of the pre- and post-test scores of both instruction groups.

In order to compare the effectiveness of the DGBL treatment, we first checked for pre-existing differences by conducting an analysis of variance (ANOVA) with pre-test as dependent and instruction method as independent variable. Results show that the DGBL group scored significantly higher on the pre-test than the group that received a slide-based lecture, F(1,71) = 20.31, p < 0.01. Two types of analyses can be distinguished in the literature which allow to take these pre-existing differences into account: an ANOVA on the change scores and an analysis of covariance (ANCOVA) with pre-test scores as covariate (Dimitrov et al., 2003; Knapp & Schafer, 2009). Considering that there is no agreement on which one to use and that the aim of the present study is to explore the disadvantages and advantages of adding a pre-test in your study design, we provide results for both.

Results of the ANCOVA show that, after controlling for initial differences on the knowledge test, instruction type still has an effect on the post-test scores showing a medium effect size (1,71) = 18.36, p < 0.01, r = 0.35. More specifically, the group receiving the DGBL treatment outperformed participants receiving the slide-based lecture on fire safety knowledge (see Table 5). Consequently, based on the ANCOVA, we can state that the DGBL treatment is more effective for the fire safety training than the slide-based lecture regarding learning outcomes. When we conduct an ANOVA on the change scores, however, we do not find a difference F(1,71) = 0.22, p = 0.88, r = 0.02. This would imply that both groups showed a similar learning gain as a result of the intervention (see Table 4).

An ANOVA on the post-test scores of the Instructional Materials Motivation Survey shows a significantly higher score for the groups who received the game-based intervention with a medium effect size, F(1,66) = 8.64, p = 0.01, r = 0.34. When looking at the subscales, a difference can be found for confidence (p < 0.01, r = 0.37), satisfaction/success (p < 0.01, r = 0.35) and for attention (p = 0.07, r = 0.25) but not for relevance (p = 0.12, r = 0.19). All participants thus perceived both the slide- and game-based method as relevant to their professional or personal context. However, the participants who received the slide-based lecture felt less satisfied with the learning experience (i.e., did not feel rewarded for it) and perceived more of an imbalance between knowledge/skills and the challenge that the instruction brought forward, compared to the game-based intervention. Moreover, participants in the DGBL conditions felt that the game succeeded more in gaining and keeping their attention during the intervention compared to participants in the lecture condition.

Regarding time management, an ANOVA on the time spent shows that the participants receiving a slide-based lecture spent a significantly higher amount of time on the intervention, with a large effect size F(1,71) = 54.61, p < 0.001, r = 0.66. More specifically, the lecture took on average 9.17 min longer. Consequently, the game-based intervention is more effective regarding efficiency outcomes.

3.1.2. Post-only design (N = 60)

When we compare the post-test data of participants who only received a post-test, an effect of treatment on performance can be detected in favor of the DGBL intervention with a large effect size, F(1,58) = 104.22, p < 0.001, r = 0.80.

Table 4
Descriptive statistics of pre-test post-test design (N = 73).

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Pre-test score (M/SD)</th>
<th>Post-test score (M/SD)</th>
<th>Minimum score (pre-test/post-test)</th>
<th>Maximum score (pre-test/post-test)</th>
<th>Adjusted post-test score (M/SD)</th>
<th>Gain score (M/SD)</th>
<th>Time spent on the intervention</th>
<th>Total score (M/SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Game group</td>
<td>34</td>
<td>16.19/7.81</td>
<td>34.44/5.64</td>
<td>3.5/17</td>
<td>33/39.5</td>
<td>33.01/184</td>
<td>18.03/5.45</td>
<td>35.08 min</td>
<td>4.21/48</td>
</tr>
<tr>
<td>Lecture group</td>
<td>39</td>
<td>9.27/5.5</td>
<td>27.29/5.31</td>
<td>1/14.5</td>
<td>18/35</td>
<td>28.50/91</td>
<td>18.25/7.30</td>
<td>25.18 min</td>
<td>3.86/48</td>
</tr>
</tbody>
</table>
Table 5
Descriptive statistics of the post-test only design.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Post-test score (M/SD)</th>
<th>Minimum score post-test</th>
<th>Maximum score Post-test</th>
<th>Time spent on the intervention</th>
<th>Total score on IMMS (M/SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Game group</td>
<td>31</td>
<td>35.05/3.78</td>
<td>25.50</td>
<td>39.5</td>
<td>35 min</td>
<td>4.14/36</td>
</tr>
<tr>
<td>Lecture group</td>
<td>29</td>
<td>21.60/6.65</td>
<td>7</td>
<td>31.5</td>
<td>24.66 min</td>
<td>3.98/41</td>
</tr>
</tbody>
</table>

No difference can be found for the IMMS, \( F(1,54) = 2.61, p = 0.11, r = 0.2 \). When we look at the subscales, however, a difference can be found for satisfaction in favor of the DGBL training, \( F(1,56) = 5.021, p = 0.01, r = 0.44 \).

Regarding time management, an ANOVA on the time spent shows that the participants receiving a slide-based lecture spent a significant amount of time more on the intervention, with a large effect size \( F(1,55) = 46.42, p < 0.001, r = 0.68 \). More specifically, the lecture took on average 10 min longer. Consequently, the game-based intervention is more effective regarding efficiency outcomes.

3.2. Effect of the pre-test

Considering that there are no guidelines for the types of analyses to conduct when pre-existing differences exist in a Solomon 4-group design, we conducted our analysis twice: once with the complete data set (i.e., including individual differences on the pre-test of knowledge) and once matching participants who received a pre-test on their pre-test scores (\( N = 102 \)).

3.2.1. Analysis on complete dataset

In order to assess the influence of the pre-test on both the post-test and the treatment, we conducted a \( 2 \times 2 \) ANOVA as suggested by Braver & Braver (1988). The two independent factors where the administration of a pre-test (two levels: pre-test was administered or no pre-test was administered) and the instruction type (two levels: DGBL or slide-based lecture). The dependent variable was post-test score. All statistics below are based on the transformed data, but the graphs reflect the untransformed data. Results show that there is a very large main effect of instruction type \( F(1,129) = 136.67, p < 0.01, r = 0.71 \). More specifically, the participants that received the DGBL intervention scored significantly higher on the post-test (see Fig. 1).

The results show a small main effect of administering a pre-test \( F(1,129) = 5.22, p = 0.02, r = 0.14 \) and an interaction between pre-test and instruction type with a small effect size \( F(1,129) = 7.46, p = 0.01, r = 0.17 \). In Fig. 2, we see that the influence of the pre-test on the treatment is larger in the group that received a slide-based lecture than in the group that received a DGBL intervention (see Fig. 3).

When we compare post-test scores of the four groups using an ANOVA with the grouping variable (four levels: DGBL with pre-test, DGBL without pre-test, slide-based with pre-test and lecture without pre-test) again we see a very large effect of instruction method, \( F(3,129) = 47.44, r = 0.72 \). A post-hoc Scheffé test shows that no differences can be found between the DGBL group that received a pre-test and the DGBL group that did not receive a pre-test (\( p = 0.99 \)). A difference is detected between the slide-based lecture group that did receive a pre-test and that which did not however (\( p < 0.01 \)). More specifically, the group that received a pre-test before receiving the slide-based lecture, scores significantly higher than the group that did not receive a pre-test before the slide-based lecture. This indicates that administering a pre-test influences the participants’ sensitivity to receiving the fire safety training with a lecture, resulting in higher scores on the post-test. This is not the case when receiving the training by DGBL.

![Fig. 1. Screenshots from the game.](image-url)
Fig. 2. Line plot of mean post-test scores ($N = 133$); dashed line corresponds to 'game' and solid line to 'lecture'.

Fig. 3. Line plot of mean post-test scores ($N = 102$); dashed line corresponds to 'game' and solid line to 'lecture'.

Both gaming groups still score significantly higher on the post-test scores compared to the lecture groups, indicating that the game is more effective in terms of knowledge transfer than the slide-based lecture.
3.2.2. Analysis on matched groups

Matched groups were constructed for the participants that received a pre-test, by looking for participants in the DGBL and the slide-based lecture group that have a similar score (i.e., maximum 1 point difference) (Rubin, 1973). In the end, 21 participants remained in both the DGBL and lecture group that received a pre-test. No differences were found on pre-test scores between the newly composed experimental and control groups receiving a pre-test, \( F(1,40) = 0.02, p = 0.82 \). Since the other groups did not receive a pre-test, we could not match them based on pre-test scores and thus left them unmodified. The present analysis was conducted on a sample of 102 participants.

In order to test for pre-test influences, we conducted the same 2 × 2 ANOVA as discussed in 3.2.1. In line with the results on the complete dataset, we find a main large effect of instruction type, in favor of the DGBL intervention, \( F(1,98) = 46.29, p < 0.01, \ r = 0.59 \). The main effect of pre-test, however, disappears \( F(1,98) = 2.7, p = 0.1, \ r = 0.12 \). An interaction between instruction type and pre-test administration is still detected, showing a small to medium effect \( F(1,98) = 16.42, p < 0.01, \ r = 0.28 \). When we look at the graph, we again see that the influence of the pre-test on the treatment is larger in the group that received a slide-based lecture than in the group that received a DGBL intervention.

When conducting an ANOVA on the post-test scores of the 4 groups, instruction has a medium to large effect on post-test scores, \( F(3,98) = 34.98, p < 0.01, r = 0.46 \). A post-hoc Scheffé test shows that differences can be found between lecture with pre-test and the lecture without pre-test groups (\( p < 0.01 \)). More specifically, the group that received a pre-test before the slide-based lecture scored significantly higher on the post-test than participants who did not receive a pre-test before the lecture. No differences are found between the game group that received a pre-test and the game group that did not receive a pre-test (\( p = 0.41 \)).

The game groups significantly outperformed both lecture groups (\( p < 0.05 \)), indicating that the game is more effective in teaching the fire safety training to the hospital personnel than the slide-based lecture in terms of learning outcomes.

3.2.3. In-game behavior

During the experiment, log data regarding in game actions were collected from participants in the game-based intervention group. During game play, players were able to retrieve ‘information cards’ regarding correct procedures. The number of times players opened these cards was logged by the system. These information cards are also opened automatically at the beginning of each training. Total time spent on studying the information cards (both opened automatically and manually) was logged. Mistakes made during the game and scores were also logged. These log data provide us with an opportunity to check whether participants receiving a pre-test before the DGBL intervention behave differently from participants who did not and thus objectively test whether or not pre-test sensitization took place. As can be seen in Table 6, no differences can be found regarding number of information cards that were consulted manually, total time spent on these information cards, in-game scores or total time spent on the intervention.

4. Discussion & conclusion

The aim of this paper was twofold. On the one hand, we studied the effectiveness of a digital game-based fire safety training compared to the traditional lecture-based type. On the other hand, we assessed the impact of the administration of a pre-test on learning outcomes by means of a Solomon four group design. We will first discuss the results of the pre-test impact in order to be able to more accurately interpret the results of the effectiveness study.

Our results revealed that the pre-test influence on an educational intervention depends on the type of instruction that is administered. More specifically, pre-test sensitization was detected among the participants receiving the more traditional slide-based lecture, but not among those receiving the DGBL intervention. Participants receiving a pre-test before receiving the slide-based lecture were thus more sensitive to the intervention and consequently scored significantly higher on the post-test than participants that did not receive a pre-test before the slide-based lecture. Providing a pre-test to participants receiving a DGBL intervention did not result in higher scores on the post-test compared to participants that did not receive a pre-test before the DGBL fire safety training. When receptivity to an intervention is altered due to the pre-test in one group and not in the group to which it is compared to, bias is introduced in the design (McCambridge, Butler-Bhavsar, Witton, & Elbourne, 2011). This is an important implication for the DGBL research field, as effectiveness studies on DGBL often show non-significant differences compared to traditional instruction (Backlund & Hendrix, 2013). In pre-test post-test designs this can lead to issues related to internal validity as post-test scores in control groups receiving traditional instruction might be

<table>
<thead>
<tr>
<th>Log data</th>
<th>Pre-test before DGBL training</th>
<th>No pre-test before DGBL training</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of information cards consulted manually</td>
<td>0.48</td>
<td>0.32</td>
<td>0.67</td>
</tr>
<tr>
<td>Total time spent on information cards</td>
<td>239.79 s</td>
<td>228.32 s</td>
<td>0.6</td>
</tr>
<tr>
<td>In game score for ‘small fire’</td>
<td>872.15</td>
<td>758.67</td>
<td>0.11</td>
</tr>
<tr>
<td>In game score for ‘Smoke’</td>
<td>5806.29</td>
<td>5476.12</td>
<td>0.41</td>
</tr>
<tr>
<td>In game score for ‘Blaze’</td>
<td>3922.15</td>
<td>3485.21</td>
<td>0.29</td>
</tr>
<tr>
<td>Total time spent on the DGBL intervention</td>
<td>1814.95 s</td>
<td>2001.67 s</td>
<td>0.52</td>
</tr>
</tbody>
</table>
significantly elevated as a result of administration of the pre-test while the scores in the DGBL treatment represent less biased scores. This non-significant difference might have been significant in favor of DGBL when no pre-test sensitization would have occurred in the traditional lecture. This makes comparison of post-test scores as a result of different instruction methods difficult in pre-test post-test designs.

The results of our effectiveness study reflect the issue discussed above. While we could find an effect of treatment in favor of the DGBL training when comparing post-test scores of participants receiving both a pre- and post-test and participants who only received a post-test (after controlling for initial differences), we could not find an effect of treatment when comparing progress of the participants receiving both a pre- and a post-test. Considering that, based on our results, the post-test scores of the participants in the lecture group are positively biased and that both game groups (those receiving a pre-test before the intervention and those not) still outperformed the lecture group receiving a pre-test, we conclude that the DGBL fire safety training is more effective regarding learning outcomes.

Regarding motivational outcomes, results were mixed. While the DGBL group scored better on the IMMS when comparing treatments of participants receiving both a pre- and post-test, this result was not present among participants only receiving a post-test. Results regarding time management in favor of DGBL could, however, be replicated among both the participants who received a pre- and post-test and among participants only receiving a post-test.

An explanation as to why pre-test sensitization takes place in the lecture group but not in the game group can be found in a combination of the motivation paradigm of entertainment education proposed by Ritterfeld and Weber (2006) and self-determination theory (Ryan & Deci, 2000). According to the motivation paradigm, DGBL is implemented to ‘seduce’ the learners by gameplay to allocate their attention to the learning content. Interactivity is one of the main characteristics of game-based learning (and e-learning in general) resulting in higher attention during the activity and consequently, deeper processing of the content (Ritterfeld et al., 2004). According to self-determination theory, higher levels of autonomy related to regulation for motivation to engage in an educational intervention leads to better performance (Ryan & Deci, 2000). Applied to our case study, this means that the pre-test initially primed both participants receiving the game training and the slide-based lecture to be more attentive due to a change of external motivation with external regulation (i.e., following the training because they are obligated) to introjected regulation (i.e., following the training as a need to prove their ability), leading to higher engagement during the training and higher scores on the post-test (Ryan & Deci, 2000). This can explain the higher post-test scores among participants receiving a pre-test before the slide-based lecture, compared to those who did not receive a pre-test. The interactivity of the fire safety game, however, requires the participants to focus their attention to the content to complete the course, regardless of whether they received a pre-test or not. This can explain the similar scores and similar in-game behavior among participants receiving a pre-test before the DGBL training and those who did not. Moreover, the interactivity of the game training possibly stimulated a shift towards a more intrinsic motivation to finish the training for both game groups (Clark, 2007), which can explain the higher scores of the DGBL groups compared to the slide-based groups. This is supported by our data which show a significantly higher score on the attention subscale of the IMMS in favor of the DGBL training when comparing treatments among participants who received both a pre- and post-test. We could not replicate this finding when comparing treatments among participants who only received a post-test. However, the researchers who were present during all interventions, perceived participants in the lecture groups who did not receive a pre-test before the intervention as more noisy and less attentive. Hence, it could be that socially desirable answers are blurring our data. A motivational finding consistent over all designs, however, was the difference between the game groups and lecture groups on the satisfaction subscale in favor of the DGBL training. This again, can be an argument in favor of a shift to intrinsic motivation in the DGBL groups, as a rewarding experience is considered a key characteristic of games for stimulating intrinsic motivation (Hwa Hsu, Lee, & Wu, 2005).

An alternative explanation as to why pre-test sensitization occurred in the slide-based group and not in the game group can be related to the fact that the pretest created curiosity for the slide-based lecture as a result of an information gap that was created by the pre-test (i.e., they did not know or only partly knew the answers to certain questions, enhancing their motivation to acquire this information during the intervention). This is called epistemic curiosity and is a known and empirically tested mechanic that can stimulate curiosity and consequently, learning (Pluck & Johnson, 2011). In the groups without a pre-test, this information gap was not created. As the game is a novel and uncommon medium to receive the fire safety training in the hospital, it could be that curiosity was also created for participants that did not receive a pre-test before the game-based intervention. Curiosity as a result of novelty called perceptual curiosity and can attract attention and consequently, enhance learning (Pluck & Johnson, 2011). However, this is not in line with a study conducted by Wouters, Oostendorp, Boonekamp, and Spek (2011), who did find an impact of creating an information gap by a foreshadowing/backstory technique on curiosity among a digital game-based learning intervention. However, no significant differences could be found regarding recall, which is similar to our results. Hence, it would have been interesting to assess curiosity regarding the learning content right before and after the interventions.

Furthermore, the effectiveness of the game-training can be explained by the constructivist learning mechanics in the game. In this case, the game provides an experiential (e.g. something happened in a hospital room), problem-based (e.g., how do you deal with smoke?), learn-by doing (e.g., show which steps you need to take to evacuate a patient from a room) and trial and error approach. Moreover, the game implements a situated learning experience by simulating the authentic environment to which the learning applies. The results of this study thus support that constructivist learning approaches provide more effective learning than passive approaches, at least for declarative and procedural learning purposes.
Regarding the effectiveness of DGBL in the present study, we can thus conclude that the added value of the game can be found in its ability to 1) attract and keep the attention of the learning during the whole course of the intervention as a result of its interactivity, 2) stimulate intrinsic motivation during the intervention, 3) stimulate curiosity during the intervention and 4) provide the constructivist learning approaches.

With the present study, we have also established the advantages of adding a pre-test, indicating pre-existing differences between experimental and control group. Consequently, when looking into the effectiveness of the DGBL treatment, we could control for these initial differences by adding pre-test scores as a covariate and thus have a more precise estimate of our treatment effect (Dimitrov et al., 2003). Bearing in mind the advantages and disadvantages of the administration of a pre-test as described above, we have several recommendations for researchers aiming to assess the effectiveness of DGBL. Firstly, pre-tests should be administered but time between pre- and post-test should be increased, minimizing the influence of the pre-test (Dochy, Mien, & Michelle, 1999). This also gives researchers the opportunity to match participants in experimental and control group, based on their pre-test scores (Gerber & Green, 2012). Secondly, we recommend researchers to not only report on differences between groups regarding progress (i.e., gain scores), but also on post-test scores, in order to provide a more complete understanding of the data, as these can yield different results (Knapp & Schafer, 2009).

On a final note, while the Solomon design seemed to be promising, there is a lack of clear guidelines on how to analyze its data. In the present study we bumped upon the issue of partly failed randomization and thus higher scores on the pre-test in the experimental group. While we aimed to solve this issue by conducting an analysis on matched groups regarding pre-test scores, we still do not have an indication as to what extent the groups that did not receive a pre-test had the same baseline knowledge regarding fire safety. The Solomon design seems to be a good design if one has perfect data, but in the social sciences, data is often not textbook perfect and researchers do not have the opportunity to gather over thousands of data points assuring successful randomization. Nevertheless, in the present study, the Solomon design has proven its added value, as it provides us with a more nuanced view of our data. For instance, we can make a more supported claim on the DGBL effectiveness regarding learning outcomes and know that we have to be careful with the interpretation of our results regarding motivational outcomes.

5. Limitations and further research

Further research implementing the Solomon four group design is required, as there were pre-existing differences between the experimental and control group in the pre-tested groups. This keeps us in doubt about the similarity of the experimental and control groups in the un-pretested groups regarding prior knowledge on fire safety, possibly influencing our results. Hence, further validation of our results is required. A limitation of the present study is that the group dynamics in the experimental and control groups are not the same (individual vs. group). However, as the aim of the hospital is to replace the traditional slide based lecture by the fire safety training game, the comparison we made is meaningful. Moreover, the game-based group received the game intervention in a controlled setting (e.g., in a hospital conference room where researchers were present), requiring participants to fully aim their attention towards the game in order to complete it. It would be interesting to investigate whether the same results would be achieved if the game is played in a less controlled setting, where the participants are asked to play the game at their own convenience within a certain timespan. Furthermore, our study took place in a professional context and participants were adults, which was also the case in the study of Arbaugh (2000) who also found pre-test sensitization effects in a traditional class, but not in an internet-based course. A test situation is not very common among this population. It would thus also be interesting to replicate a similar study among a student population, where tests and exams are more common.

In the present study, we have only focused on the effect of a pre-test in terms of cognitive learning outcomes. More specifically, we have focused on the effects of administering a pre-test of knowledge on the learning outcomes which consist of the same knowledge assessed pre-intervention. Considering that our motivational outcomes concerned an evaluation of the instructional material, this could only be administered post-intervention. An interesting venue for further research is to study the impact of administering a pre-test regarding motivational outcomes that are assessed pre- and post-intervention, such as interest for the learning content or attitudes towards a certain subject.

In hindsight, it would also have been interesting to add observational data on the control groups in our study to get insight into whether participants receiving a pre-test before the slide-based lectures actually behave differently than those that did not receive a pretest before the lecture as an indicator for pre-test sensitization. The researcher present during the intervention had the impression that participants receiving a pre-test before the slide-based lecture were more attentive than participants in the second control group, but these are subjective perceptions carrying only limited validity.

Finally, a follow-up study to see longer term effects of both interventions and the longer term effects of the pre-test would be of added value for the present study and will be.

Acknowledgements

The current research paper is part of a Ph.D. project which is funded by the IWT, the Flemish government agency for Innovation by Science and Technology (IWT). The sponsor's role is limited to an evaluation after the second year of the Ph.D. project. Consequently, the sponsor is not involved in the study design, data collection, analysis and interpretation of data; in
the writing of the report; and in the decision to submit the article for publication. The game that has been tested in the present study, was funded by the European Social Fund. They were not involved in the effectiveness study.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.compedu.2017.05.018.

References


