Pre-test session impact on the effectiveness assessment of a fire safety game

Anissa All[1], Barbara Plovie[2], Elena Patricia Nuñez Castellar[1,3] & Jan Van Looy[1]
[1]iMinds, MICT, Ghent University, Ghent, Belgium
[2] University college, West-Flanders
[3]Department of Data Analysis, Ghent University, Belgium
Anissa.All@Ugent.be
Barbara.Plovie@howest.be
ElenaPatricia.NunezCastellar@Ugent.be
J.Vanlooy@Ugent.be

In recent years, critiques have been formulated regarding current evaluation methods of DGBL (digital game-based learning) effectiveness, putting the validity of certain results in doubt. An important point of discussion in DGBL effectiveness studies is whether or not a pre-test should be administered, as it can lead to practice effects and pre-test sensitization, threatening internal validity of the results.

The present study aims at testing if the administration of a pre-test has a direct influence on post-test scores and/or makes participants more receptive to the intervention. 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 34 participants received a pre-test and 31 did not. The control groups received traditional classroom instruction (n=68), of which 39 participants received a pre-test and 29 did not. A 2x2 ANOVA was used to explore the practice effect and the interaction between the pre-test and the intervention. An interaction effect between pre-test and intervention is detected. More specifically, this interaction takes place in the traditional classroom group, indicating pre-test sensitization.

In the traditional classroom context, the pre-test makes the participants more sensitive to the content treated in the intervention while administration of a pre-test does not influence outcomes of the DGBL treatment. When the administration of a pre-test influences the control group's receptivity to the treatment, but not the experimental group, results of an effectiveness study may be biased. This is especially relevant in the DGBL field as often, non-significant differences between DGBL and more traditional methods are reported. Therefore, further research should take this into account and look for possible solutions to solve this discrepancy.

Keywords: DGBL, pre-test sensitization, practice effect, Solomon 4-group design, fire safety

1. Intoduction

The interest in using digital games as instructional tools for digital game-based learning (DGBL) has increased over the past decade. DGBL has been implemented in different sectors, such as defense, education, corporate training, health and wellbeing, and communication (Backlund & Hendrix, 2013). Whilst in recent years, there has been a significant increase in publications on the effectiveness of DGBL (Hwang & Wu, 2012), certain methodological issues remain a matter of debate in this field (Bellotti et al. 2013; Giessen 2015).

For instance, there is a large heterogeneity in study designs which makes the comparison of results across DGBL studies difficult (All, Nuñez Castellar & Van Looy). By heterogeneity we mean that research designs differ on several respects, such as implementation of a control group, activities implemented in the control group(s) or the administration of a pre-test (Hays, 2005; Michael & Chen, 2005). Furthermore, 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 often fail to mention whether or not self-developed tests have been piloted, which leads to doubts with regard to the reliability and validity of results (Brom, Šisler, Buchtová, Klement, & Levčík, 2012). Furthermore, another important methodological issue that deserves attention is that it is difficult to replicate published DGBL effectiveness studies given that authors often do not provide enough information on how the intervention - both the experimental and control condition- has been implemented (Authors). Detailed information on procedure is indispensable, however, in order to assess 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).

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 research designs is required.

1.1. Pre-test administration

An important point of discussion is whether or not a pre-test, gauging for baseline measures of knowledge, should be administered. The addition of a pre-test to the research design is advantageous as it allows researchers to control for pre-existing differences between the experimental (game-intervention) and control group (traditional intervention) (Clark 2007) and to compare progress (i.e., gain scores) as a result of the intervention implemented (Gerber and Green 2012). Moreover, a more precise estimate of the treatment effect is allowed by adding the pre-test as a covariate, controlling for individual differences on the pre-test scores (Jamieson 2004; Knapp and Schafer 2009). Lastly, the addition of a pre-test allows the researcher to control for characteristics of drop-outs (Authors). However, adding a pre-test can also 'blur' the real effect of the treatment. Firstly, administering a pre-test can result in 'practice effect', meaning that subjects that take the same test twice, do automatically better the second time, even if the intervention would not have taken place (Crawford et al. 1989). This is thus the main effect of the pre-test, as it can offer participants additional exercise material, item training or a search strategy (van Engelenburg 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 and Braver 1988; van Engelenburg 1999). This means that subjects who have received a pre-test will be more sensitive to the intervention compared to subjects who have not received a pre-test, resulting in higher scores on the post-test. Hence, one cannot know whether a positive effect as a result of the treatment would have been present if a pre-test was not administered. Hence, generalization of results from a pre-tested to an unpretested sample is prevented. This has resulted in researchers renouncing a pre-test when studying effectiveness of DGBL (Amory 2010; Tsai et al. 2012). However, pre-test influences have -to our knowledge- never been studied in a DGBL context. Therefore, before making assumptions on the absence of a pre-test effect or pretest sensitization, this needs to be studied (Braver and 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). In this design, four conditions are present: the first two conditions are the same as in the classic pretest-posttest design: participants receive a pretest, an intervention is implemented and a post-test is administered. The two extra conditions parallel the treatment and control condition, but a pre-test is absent (see table 1).

Table 1: Solomon four group design

Condition	Pre-test	Intervention	Post-test
Treatment condition 1	Yes (O1)	X	Yes (O2)
Control condition 1	Yes (O3)	C	Yes (04)
Treatment condition 2	No	X	Yes (05)
Control condition 2	No	C	Yes (06)

The present study aims at looking at pre-test influences in a DGBL effectiveness research context. 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).

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 PowerPoint lecture. Randomization of subjects was not possible in this study, as the traditional lecture takes place once a month and staff already subscribed for these courses. Consequently, the groups that were formed as a result of these subscriptions were randomly assigned to either the condition with or without pre-test. During the period of the intervention, the prevention manager of the hospital organized 'extra safety training sessions' for which hospital staff could subscribe. These were also randomly assigned on a group level (i.e., a group was composed of people that subscribed for a safety training on the same date; similar to how groups are composed in the traditional lecture groups) to either the game condition with or without a pre-test.

2.2. Stimulus material

2.2.1. Digital game-based fire safety training

The DGBL fire safety training was specially developed for the hospital of which personnel participated in the study. All hospital personnel (i.e., doctors, nurses, cleaning personnel, administrative staff, technical staff, etc.) is required to follow the fire safety training every year. However, because the hospital has expanded over the years and is still expanding -a fourth campus has recently been built- organizationally, it is becoming more difficult to provide everyone this training. Hence, the decision to develop a digital game in cooperation with DAE research at the applied University of West-Flanders. The game consists of three minigames or courses: 'small fire'; 'smoke' and 'blaze'. After these courses are completed, one can also play a random 'fire safety' scenario, where elements learned in the course can be practiced. In total, 6 different scenarios are available. The game is freely available on the following website: http://sggo.howest.be/het-serious-game/

2.2.2. PowerPoint

The PowerPoint lecture is instructed by either the prevention manager or another fixed employee working at the department prevention. This is the lecture that is currently being used as a fire safety training for the hospital personnel. This lecture was also used as a base to define content treated in the game and contains exactly the same content as treated in the game.

2.3. Procedure

2.3.1. Experimental groups

The experimental groups played the game in a conference room in one of the four campuses of the hospital, during their working hours. A maximum of six subjects could participate per session. When entering the conference room, subjects received an introduction with information 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 help was only provided if there were issues with the computer or game play. After the subjects completed all three courses and one scenario, a pre-test was administered. A maximum of 6 participants could take part in a game based fire safety training session. 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 received the PowerPoint lecture in a conference room in one of the four campuses. The PowerPoint lecture was instructed 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 the 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 PowerPoint lectures were organized, 3 included a pre-test and 3 did not.

2.4. Participants

The present study was in collaboration with a hospital in <location removed for blind review process>. In total, 152 subjects participated in the study. Eighty tree subjects participated in the experimental groups, of which 42 subjects received a pre-test and 41 did not receive a pre-test. Sixty nine subjects participated in the control groups of which 39 received a pre-test and 30 did not receive a pre-test. Eightteen subjects in the experimental group (8 that received a pre-test and 10 that 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. In the end, 134 participants were included in the analysis.

As can be seen in table 2 provides, 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 proportion.

Table 2: Control for balanced groups as a result of randomization on group level

	Experimental group with pre-test (n=34)	**Experimental group without pre-test** (n=31)	**Control group with pre-test** (n=39)	**Control group without pre-test** (n=29)	**Chi²/F**	**p**
Female gender	76.50%	71.00%	92.30%	96.60%	10.87	**0.01**
Age (mean)	40.03	37.52	38.31	40.83	0.54	0.66
Gamers	50.00%	61.30%	61.50%	48.10%	2.00	0.57

2.5. Measures

2.5.1. Cognitive learning outcomes

In order to assess knowledge a test was developed by the researchers in cooperation with the prevention staff responsible for the fire safety training –the same staff that provides the PowerPoint lectures. The test was previously implemented in a pilot study in an initial phase of development of the game. The test consists of open ended questions, covering all learning elements that are treated in both interventions. The test consists of 18 questions with a maximum score of 40. 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 to follow when evacuating patients? Which tree 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 doubt on the correctness of certain answers, researchers consulted each other to agree upon a score.

2.5.2. Motivational outcomes

The IMMS -Intructional Materials Motivation Survey- (Keller 1987) was used to assess motivation towards the instruction method. We based ourselves on (Huang et al. 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 are scored on a 5-point Likert scale. The total score represents motivation towards the instructional material and scores on the subscales, give an indication on which elements the instruction failed, based on the subcomponents (Keller 2010). Enjoyment was assessed by using Ryan & Deci's (Ryan 1982) enjoyment/interest scale from the post-experimental intrinsic motivation inventory. The scale consists of 8 items which are scored on a are scored on a 7-point Likert scale. Interest was assessed by the social facilitation subscale of the subjective involvement scale developed by This was assessed by a three item-scale developed by Neys & Jansz (2010). This scale aims at assessing the desire to interact with others about the topic discussed in the game and consists of 3 items. The items were scored on a 7-point Likert scale. We, however, do not discuss the motivational outcomes of the present study, as space is limited and the aim of the paper is to assess the impact of the pre-test and motivational outcomes -in this case- can only be assessed post-intervention.

3. Results

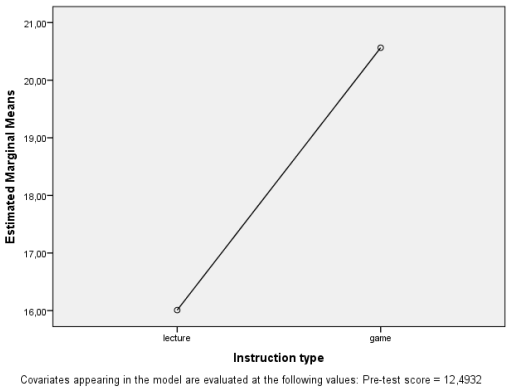
3.1. Effectiveness of the DGBL treatment

We firstly conducted an analysis on the groups receiving a pre-test (N=73), in order to determine progress (pre-test vs. post-test scores) and compare progress between groups (DGBL vs. PowerPoint lecture). A paired sampled t-test showed a significant difference between pre-test and post-test scores for both the participants receiving a PowerPoint lecture, t(33)= 9.45, p=0.00 and the participants in the DGBL intervention t(38)=12.78, p=0.00, showing that both the DGBL training as the PowerPoint lecture entail a learning effect. Since the post-test scores were not normally distributed, we conducted the analysis on transformed data (square root transformation)(Kutner et al. 2005).

In order to compare whether or not this progress is different across groups, we first checked for pre-existing differences by conducting an ANOVA with pre-test as dependent and instruction method as independent variable. Results show that the DGBL group scores significantly higher on the pre-test than the group that received a PowerPoint lecture, F(1,71)=20.31, p=0.00. Hence, we need to take these pre-existing differences in our analysis by adding pre-test scores as covariates (Jamieson 2004).

We used a difference-in-difference approach for our data-analysis (Gerber and Green 2012), creating a new variable 'gainscore' by subtracting the pre-test scores from the post-test scores. An ANCOVA was conducted with gainscore as dependent variable, instruction type as independent variable and pre-test scores as a covariate, to control for initial differences (Jamieson 2004). Results show that after controlling for initial differences, instruction type has a significant effect on the game scores. In figure 1 (reflection of untransformed data), we see that the DGBL group shows significantly higher gain than the group that received a PowerPoint lecture, F(1,79)=10,76, p=0.00. Note that if we would not have controlled for these initial differences on pre-test, no effect of instruction type would have been found, F(1,79)=0.02, p=0.88.

Fig. 1: Line plot of mean gain score



3.2. Effect of the pre-test

Considering that there are no clear guidelines on the types of analyses to conduct when implementing a Solomon 4-group design when pre-existing differences exist between experimental and control group that received a pre-test, we will conduct our analysis twice: one with the complete data set (i.e., including individual differences) and one where we have matched participants that received a pre-test on their pre-test scores (N= 102)

3.2.1. Analysis on complete dataset

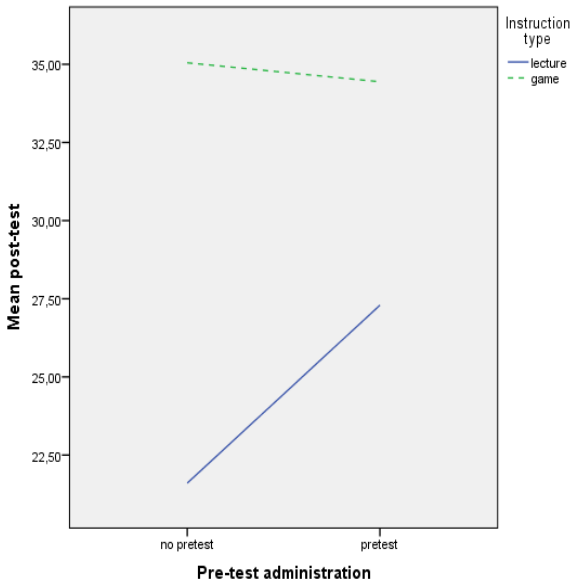
In order to assess the influence of the pre-test on both the post-test and the treatment, we conducted a 2x2 ANOVA. The two independent factors where the administration of a pre-test (two levels: pre-test was administered or no pre-test was administered) and the instruction type (two levels: DGBL or PowerPoint lecture). The dependent variable was post-test scores. Considering that the assumption on normality was violated and the data were negatively skewed, we conducted the analysis on transformed data (square root transformation) (Kutner et al. 2005). All statistics below are based on the transformed data, but the graphs reflect the untransformed data.

Results show that there is a main effect of instruction type $F(3,147)=10.19$, $p=0.00$. More specifically, the participants that received the DGBL intervention scored significantly higher on the post-test.

There is no main effect of administrating a pre-test on the post-test scores $F(3,15)=3.44$, $p=0.06$, but that the interaction between pre-test and instruction type is significant $F(3,15)=10.19$, $p=0.00$. In fig. 2, we see that the influence of the pre-test on the treatment is larger in the group that received a PowerPoint lecture than in the group that received a DGBL intervention.

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, PowerPoint Lecture with pre-test and PowerPoint Lecture without pre-test), a post-hoc Scheffé test shows that no significant differences can be found between the DGBL intervention group that received a pre-test and the DGBL intervention group that did not receive a pre-test, $F(3,13)=47.44$, $p=0.99$. A significant difference is detected between the PowerPoint lecture groups that did receive a pre-test and those that did not ($p=0.00$). More specifically, the group that received a pre-test before receiving the PowerPoint lecture, scores significantly higher than the group that did not receive a pre-test before the PowerPoint lecture. This indicates that administering a pre-test influences the participants' sensitives to receiving the fire safety training with a PowerPoint Lecture, but not when receiving the training by playing the game.

Fig. 2: Line plot of mean post-test scores (N=134)



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

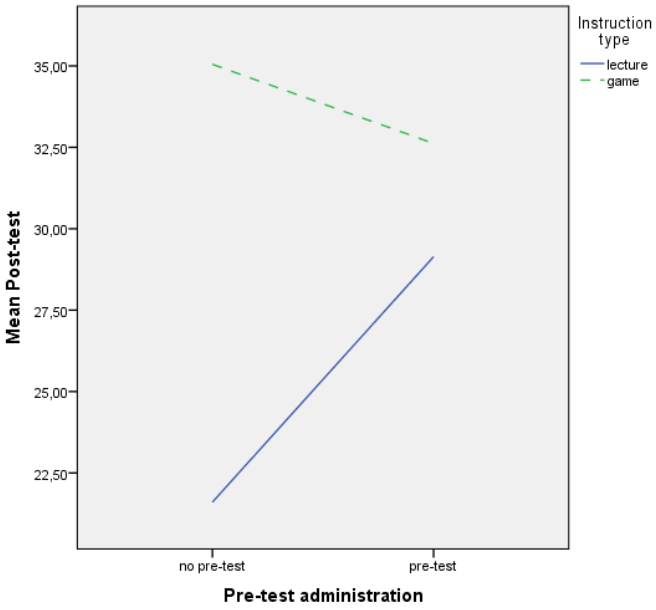
3.2.2. Analysis on matched groups

Matched groups were constructed for the participants that received a pre-test, by looking for participants in the DGBL and the PowerPoint lecture group that have a similar score (i.e., maximum 1 point difference). In the end, 21 participants remained in both the DGBL and PowerPoint group that received a pre-test. No significant differences were found on pre-test scores between the new composed experimental an 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 2X2 ANOVA as discussed in 3.2.1. In line with the results on the complete dataset, we find a significant main effect of instruction type, in favor of the DGBL intervention, F(3,98)=69.33, p=0.00. No significant main effect of pre-test is found, F(3,98)=2.97, p = 0.09 and a significant interaction between instruction type and pre-test administration is detected, F(3,98) =16.99, p=0.00. When we take a 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 PowerPoint lecture than in the group that received a DGBL intervention.

When conducting an ANOVA on the post-test scores of the 4 groups, a post-hoc Scheffé test shows that there is no significant difference between the post-test scores in the DGBL training group between participants that received a pre-test and those that did not F(3,98)=34.61, p=0.41. A significant difference can, however, be found between groups receiving PowerPoint instruction that were administered a pre-test before the lecture and those that were not. More specifically, the group that received a pre-test before the PowerPoint lecture scored significantly higher on the post-test than participants that did not receive a pre-test before the PowerPoint Lecture.

Fig. 3: Line plot of mean post-test scores (N=102)



The game groups outperformed both PowerPoint groups, indicating that the game is more effective in teaching the fire safety training to the hospital personnel than the PowerPoint lecture.

4. Discussion and conclusion

The present study has for the first time -to our knowledge- conducted a Solomon four group design in the context of DGBL. 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 found when receiving the traditional PowerPoint lecture, but not when receiving the DGBL intervention. Participants receiving a pre-test before receiving a PowerPoint lecture on fire safety 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 PowerPoint lecture. Providing a pre-test to participants receiving 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 et al. 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 (Giessen 2015). In pre-test post-test designs this can lead to issues regarding internal validity, as post-test scores in control groups receiving traditional instruction might be significantly elevated by the administration of the pre-test while the scores in the DGBL treatment represent the 'true' scores as a result of the instruction itself. This non-significant difference might have been significant in favor of DGBL when no pre-test sensitization would have occurred in traditional lecture. This makes comparison of post-test scores as a result of different instruction methods rather difficult.

In the present study, the game groups still outperformed both the control group receiving a pre-test and the control that did not receive a pre-test, concluding that the DGBL fire safety training was highly effective.

The present study has also shown the advantages of adding a pre-test, indicating pre-existing differences between experimental and control group. This way, when looking into the effectiveness of the DGBL treatment, we could control for these initial differences by adding pre-test scores as a covariate (Jamieson 2004; Knapp and Schafer 2009). If we would not have been able to control for initial differences, no significant differences would have been found when comparing the progress of the PowerPoint lecture and the DGBL group.

Considering the advantages and disadvantages of the administration of a pre-test as described above, we have several recommendations for researchers aiming at assessing 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 et al. 1999). This also gives researchers the opportunity to match participants in experimental and control group, based on the pre-test scores (Gerber and Green 2012). Secondly, we recommend using parallel tests pre- and post- interventions (i.e., same types of questions and same difficulty level). These tests should be piloted beforehand, in order to check whether or not these tests can be perceived as parallel versions. A good example can be found in (Nunez Castellar et al. 2013). Thirdly, 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 and Schafer 2009).

5. Limitations and Further research

Further research implementing the Solomon design is required, as there were pre-existing difference 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 unpretested groups regarding prior knowledge on fire safety, possibly influencing our results. Hence, further validation of our results is required.

6. References

Amory, Alan. 2010. "Learning to play games or playing games to learn? A health education case study with Soweto teenagers." Australasian Journal of Educational Technology 26(6):810-829.

Bellotti, Francesco, Bill Kapralos, Kiju Lee, Pablo Moreno-Ger and Riccardo Berta. 2013. "Assessment in and of serious games: an overview." Advances in Human-Computer Interaction 2013:1.

Braver, Mary W and Sanford L Braver. 1988. "Statistical treatment of the Solomon four-group design: A meta-analytic approach." Psychological Bulletin 104(1):150.

Clark. 2007. "Learning from serious games? Arguments, evidence, and research suggestions." Educational Technology 47(3):56-59.

Crawford, JR, LE Stewart and JW Moore. 1989. "Demonstration of savings on the AVLT and development of a parallel form." Journal of Clinical and Experimental Neuropsychology 11(6):975-981.

Dochy, Filip, Mien Segers and Michelle M Buehl. 1999. "The relation between assessment practices and outcomes of studies: The case of research on prior knowledge." Review of Educational Research 69(2):145-186.

Gerber, A.S. and D.P. Green. 2012. Field Experiments. Design, Analysis and Interpretation. : W. W. Norton & Company.

Giessen, Hans W. 2015. "Serious Games Effects: An Overview." Procedia-Social and Behavioral Sciences 174:2240-2244.

Huang, Wen-Hao, Wen-Yeh Huang and Jill Tschopp. 2010. "Sustaining iterative game playing processes in DGBL: The relationship between motivational processing and outcome processing." Computers & Education 55(2):789-797.

Jamieson, John. 2004. "Analysis of covariance (ANCOVA) with difference scores." International Journal of Psychophysiology 52(3):277-283.

Keller, John M. 1987. "Development and use of the ARCS model of instructional design." Journal of instructional development 10(3):2-10.

Keller, John M. 2010. Motivational design for learning and performance: Springer.

Knapp, Thomas R and William D Schafer. 2009. "From Gain Score t to ANCOVA F (and vice versa)." Practical Assessment, Research & Evaluation 14(6):1-7.

Kutner, Michael H, Christopher J Nachtsheim, John Neter and William Li. 2005. "Applied linear statistical models."

Mayer, Igor, Geertje Bekebrede, Casper Harteveld, Harald Warmelink, Qiqi Zhou, Theo Ruijven, Julia Lo, Rens Kortmann and Ivo Wenzler. 2014. "The research and evaluation of serious games: Toward a comprehensive methodology." British Journal of Educational Technology 45(3):502-527.

McCambridge, Jim, Kaanan Butor-Bhavsar, John Witton and Diana Elbourne. 2011. "Can research assessments themselves cause bias in behaviour change trials? A systematic review of evidence from Solomon 4-group studies." PLoS One 6(10):e25223.

Neys, Joyce and Jeroen Jansz. 2010. "Political Internet games: Engaging an audience." European Journal of Communication 25(3):227-241.

Nunez Castellar, Elena , Jan Van Looy, Arnaud Szmalec and Lieven De Marez. 2013. "Improving arithmetic skills through gameplay: assessment of the effectiveness of an educational game in terms of cognitive and affective learning outcomes." Information sciences In Press.
Ryan, Richard M. 1982. "Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory." Journal of personality and social psychology 43(3):450.
Solomon, Richard L. 1949. "An extension of control group design." Psychological Bulletin 46(2):137.
Tsai, Fu-Hsing, Kuang-Chao Yu and Hsien-Sheng Hsiao. 2012. "Exploring the Factors Influencing Learning Effectiveness in Digital Game-based Learning." Educational Technology & Society 15(3):240-250.
van Engelenburg, Gijsbert. 1999. "Statistical Analysis for the Solomon Four-Group Design. Research Report 99-06."