



This is a repository copy of *Measuring the learning effectiveness of serious gaming for training of complex manufacturing tasks*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/124858/>

Version: Accepted Version

Article:

Li, K., Hall, M., Bermell-Garcia, P. et al. (3 more authors) (2017) Measuring the learning effectiveness of serious gaming for training of complex manufacturing tasks. *Simulation and Gaming*, 48 (6). pp. 770-790. ISSN 1046-8781

<https://doi.org/10.1177/1046878117739929>

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Measuring the Learning Effectiveness of Serious Gaming for Training of Complex Manufacturing Tasks

ABSTRACT

Background: Training new workers on complex **manufacturing tasks** has long been a challenge for high value manufacturing companies. Equipment downtime, costly instructors, and dangerous working environments are some of the impediments of hands-on training. To overcome these hurdles, a traditional manufacturing paper manual was transformed into a **serious game** through capturing and embedding expert knowledge.

Aim: This article investigates the learning effectiveness of learning via a **serious game** (Training Game) compared with the tradition learning method (Paper Manual) through a **user study**.

Method: Twenty employees took part in a randomised **controlled trial**. They were assigned to one of two conditions: Training Game (experimental condition), or Paper Manual (control condition). Participants spent a maximum of 30 minutes to study manufacturing instructions before completing two tests to evaluate the amount of learning achieved.

Results: The results show that the Training Game was more effective for learning **procedural knowledge** than the Paper Manual. Regarding **factual knowledge**, no significant difference was identified between the two conditions. In terms of motivation, increased engagement levels were reported in the Training Game condition.

Conclusions: This **user study** shows evidence that the serious TG being evaluated is an effective method for training **procedural knowledge** in a complex manufacturing scenario.

Keywords: adult learning, user study, procedural knowledge, factual knowledge, manufacturing tasks, controlled trial, computer game, training, professionals

Abbreviations: Serious Game (SG), Training Game (TG), Paper Manual (PM), Computer Aided Three-Dimensional Interactive Application (CATIA), Personal Protective Equipment (PPE), Cost of Poor Quality (COPQ)

1. INTRODUCTION

Training new workers in complex manufacturing tasks has long been a challenge for high value manufacturing companies (Mital et al., 1999). In particular training that involves hands-on experiences have associated impediments, such as the limited availability of physical equipment (Bal, 2012), health and safety concerns in the operation of dangerous equipment (Sun & Tsai, 2012), and the high training costs involved having to dedicate both the equipment and experienced professionals to instruct new workers. Serious Game (SG) applications have been studied as alternative learning methods for reducing the impact of these hurdles. SG applications are designed to be competitive and entertaining, but on top of that they contain pedagogic elements that are intended to impart knowledge to the player (Gredler, 2004; Zyda, 2005). These applications have many prospective benefits when applied to the context of complex manufacturing training. Firstly, they have the potential to improve the quality of training by providing a standardised work process to train or re-train operators (Vizendo,

2014). Secondly, they enable concurrent learning, thus reducing the training time and instructor load (Mitchell & Savill-Smith, 2004). Furthermore, SG offer the opportunity for users to train in an environment where they can learn and explore without severe consequences (Aziz, Chang, Esche, & Chassapis, 2014; Ekanayake, Backlund, Ziemke, Ramberg, & Hewagamage, 2010). Not only does this allow users to practice and rehearse situations which may be difficult or dangerous to reproduce in real world scenarios (González-Franco, Peck, Rodríguez-Fornells, & Slater, 2014), but it creates the possibility to influence a user's behaviour in real life (Mitgutsch & Alvarado, 2012; Peck, Seinfeld, Aglioti, & Slater, 2013).

SG applications are already being developed for training in a variety of disciplines, including health and safety in the construction industry (Dickinson, Woodard, Canas, Ahamed, & Lockston, 2011; Greuter et al., 2012; Kang & Jain, 2011; Mar & Zheng, 2011), medical training (Bartoli et al., 2012; Gonzalez-Franco, Gilroy, & Moore, 2014; Knight et al., 2010; Torrente et al., 2014), operations management (Lewis & Maylor, 2007), and fire services (Williams-Bell, Murphy, Kapralos, Hogue, & Weckman, 2015). However, to date, few examples of their use in complex manufacturing exists. Rare examples include the creation of virtual machines and equipment to allow learners explore at their own pace (Ong & Mannan, 2004) and do so from different localities (Bal, 2012), to provide procedural training for car maintenance (Borsci, Lawson, & Broome, 2015) and aerospace servicing (Abate, Guida, Leoncini, Nappi, & Ricciardi, 2009), and to ensure safety in the training on injection moulding machines (Sun & Tsai, 2012).

Current literature shows disputes regarding the effectiveness of SG. Even though meta-analytic studies have found gaming to be an effective instructional tool (Sitzmann, 2011; Wouters, Nimwegen, Oostendorp, Spek, & Games, 2013), it is argued that games included in the analysis lacks of rigorous testing (All, Nuñez Castellar, & Looy, 2014; Bellotti, Kapralos, Lee, Moreno-Ger, & Berta, 2013; T. M. Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012). Furthermore, there can be a huge variation between one SG to another based on the the skills being trained (Girard, Ecalle, & Magnan, 2013).

For a serious game to be considered as effective, it must accomplish what it sets out to do in a real-world context (All et al., 2015). In developing a framework for measurement the effectiveness of SG, All et al. (2015) identified three categories of assessment: learning outcome (increasing interest in the subject matter, improving objective performance, transferring knowledge or skills in real-world situations), motivational outcome (evoking enjoyment in the learning process, increasing motivation to learn with the serious game), and efficiency outcome (time management and the cost effectiveness) .

In this article, the authors intend to measure the effectiveness of a serious game designed for manufacturing training in terms of the learning outcome. The authors will hereby refer to this as learning effectiveness. There have been some attempts in categorising the different learning outcomes of games on: Cognitive Learning, Motor Skill, Affective Learning or Communicative Learning (Wouters, van der Spek, & van Oostendorp, 2009), but also on: Cognitive Outcome, Affective Outcome, Skill Based Learning Outcome (Garris, Ahlers, & Driskell, 2002), and on: Content Understanding, Problem Solving, Collaborative/teamwork, Communication, Self-Regulation (O'Neil, Wainess, & Baker, 2005). Even though numerous ways for categorising learning outcomes are available, it seems appropriate to use standardised metrics to really understand the learning effects of a SG. Similar to Buchanan et al. (2011), this study has made use of the revised Bloom's Taxonomy (Anderson & Krathwohl, 2001; Krathwohl, 2002) to describe the type of learning the TG is designed to impart. By clearly documenting the learning objectives the authors hope this study can be compared like-for-like with similar work in the future. Two learning objectives are measured in this

study. The first is for participants to recall the necessary tools, equipment, and documentations for a specific manufacturing task (Objective 1), and the second is to correctly perform an assembly operation (Objective 2). Table 1 shows how the two learning objectives are categorised.

Objective 1 is categorised in A1 (Factual-Remember). This objective requires the learner to be able to remember factual knowledge. Remember is at the lower end in the Cognitive Processes Dimension. Objective 2 is categorised in C3 (Procedural-Apply). Objective 2 requires the learner to apply the knowledge they have learnt. Applying knowledge is a higher order thinking skill compared with remembering (as in Objective 1).

TABLE 1. A Table to Show the *Learning Objectives Mapped to Bloom's Taxonomy*

The Knowledge Dimension	The Cognitive Process Dimension					
	1. Remember	2. Understand	3. Apply	4. Analyse	5. Evaluate	6. Create
A. Factual	Objective 1	-	-	-	-	-
B. Conceptual	-	-	-	-	-	-
C. Procedural	-	-	Objective 2	-	-	-
D. Metacognitive	-	-	-	-	-	-

2. MATERIALS AND METHODS

This article presents a study evaluating a serious Training Game (TG) developed to support complex manufacturing training. The TG proposed is developed from a manufacturing manual of an aircraft maintenance door. Manufacturing of civil aircrafts is subject to strict procedures due to the legal and safety implications of non-conformities. These include the design and the costs associated to the auditing and certification of the product. Through interacting with the TG, new operators are expected to achieve a reasonable level of knowledge of the assembly procedure before they are exposed to an instructor or physical manufacturing equipment. In this context the ultimate goal of the TG is to reduce the cost of poor quality (COPQ) via a more interactive and cost-effective approach to minimize product defects or deviations from the design during production. The experimental protocol employed in the present study was approved by the ethical committee of Cranfield University and the experimental data were collected with the approval and written consent of each participant.

2.1. Experimental Conditions

As a first evaluation of the Training Game, a user study was conducted to assess the learning effectiveness of the TG. This randomised controlled trial employed a post-test only control group design. The treatment group were trained on the TG, whereas the controlled group was trained on the Paper Manual (PM). This study was designed to evaluate if the TG is successful in training the assembly procedure under the condition that no instructor or physical equipment is present. Both learning materials contained the same written manufacturing instructions. The differences between the two instructional approaches are detailed in the following subsections.

2.1.1. Paper Manual

The PM used for this experiment was an original aircraft door assembly manual that is currently in use. Due to confidentiality, the PM will not be presented. It contains two main parts: 1) manufacturing instructions, and 2) supporting documents. For the first part, the manufacturing instructions are provided as a series of tasks combined with images to show how the tasks are performed. For example, one task involved removing pin bolts from the assembly jig. An image was provided to show the location of the pin bolts. For the second part, a number of support documents are presented. They

provide additional information that is not detailed in the manufacturing instruction. This included the material and tool lists, and lists of documents (such as regulations) that workers will need to refer to within the manufacturing process.

2.1.2. Training Game

The TG solution was developed in Unity3D and the implementation did not include the players' body representation, i.e. the whole TG was based on a point and click navigation within a desktop computer. To increase the fidelity of the TG, the maintenance door and assembly jig models were low-poly definitions of the original CATIA (Computer Aided Three-Dimensional Interactive Application) models; where participants could interact with all moving elements. For example, participants could rotate the jig, adjust drilling templates, or take in and out pin bolts. In essence, the TG model consists of the same functionality as a real factory jig. The following sub-sections details the features of the TG. Textual information in the images has been obscured for confidentiality purposes.

The game begins in the corridor (Figure 1) in which players can select to go into different rooms, including:

- Personal Protective Equipment (PPE) room for collecting the required protective equipment.
- Materials room for collecting materials.
- Tools room for collecting tools.
- Inspection room for inspecting tools.
- Mixing room for mixing the required solutions.
- Workshop for assembling the aircraft door.



FIGURE 1. The TG consists of a number of rooms with different functionalities, players can enter by clicking on the door

Manufacturing instructions are detailed on the left menu bar (Figure 2). These instructions are exactly the same as those in the PM. After the user has completed an operation, they would click next to move on to the next instruction. A player can view the work instructions, but they would not be scored on their performance until the 'Next Task' button is selected.



FIGURE 2. To play the game, players would follow the instructions on the left menu bar (shown in the top red box), once they have completed the required tasks they can move on to the next task by clicking the 'Next Task' button (shown in the bottom red box)

Supporting documents are provided in two different formats. First, lists of PPE, materials, and tools are located on the bottom menu bar (Figure 3). These are information that would have been provided within the PM. Secondly, additional information for regulations and procedures are selected using a 'computer' located in the workshop room (Figure 4). Currently, as in the PM, this feature only lists names of the supporting documents. This feature is intended to be further developed in later versions of the game so that on selection, the information of the supporting document is provided.



FIGURE 3. Supporting documents for the required materials and tools for each operation are provided on the bottom task bar (shown in bottom red box), once this is clicked on the list of required materials or tools are provided (shown in top red box)



FIGURE 4. Additional information of regulations and procedures are selected using a ‘computer’ located in the workshop room

The bottom menu bar (Figure 5) consists of 4 icons: material box, tool box, PPE, and a mannequin. Players can select PPE, tools and materials from the appropriate rooms and drag them accordingly to associate icon on the bottom menu bar. The mannequin represents the player. A player must equip the mannequin with correct PPE and the required tools before they can perform a task. If a player requires help, they can select the ‘Help’ button (Figure 6). Hints are provided to guide the users to complete the associated task. For example, arrows may appear to direct the players to components they are required to interact with.



FIGURE 5. PPE, tools, and materials are collected and stored on the bottom menu bar (shown in bottom red box), the mannequin represent the player, the top red box shows the figure that appears for players to be equipped with PPE



FIGURE 6. Help is provided to guide players through the game, help is given when players select the ‘Help’ button (shown in red box), for this particular task red arrows can be seen above the green component to guide the player to the correct location in which the task takes place

The health bar is located in the top right hand corner (Figure 7). It provides feedback on how well a player is doing. The amount of health would fluctuate depending on whether the player is performing a task correctly (health is added) or not (health is deducted). The game ends if no health remains. Once the game ends, players can view their total score and health status. Details of how they performed on each task are presented. The time stamp provides feedback on how long they took to complete each operation.

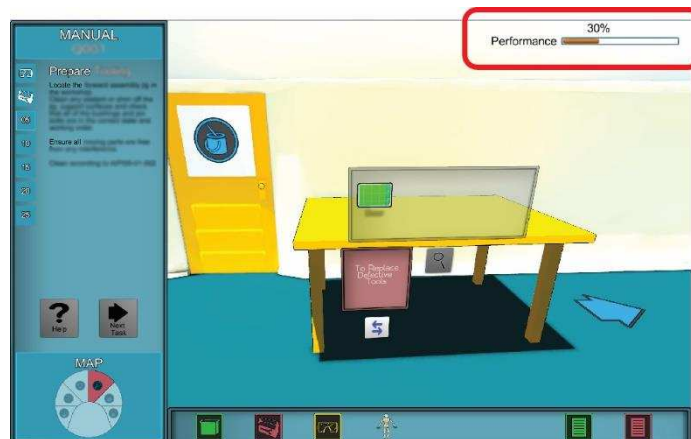


FIGURE 7. The health bar (shown in red box) is provided to inform players of their performance, the health increases if a task is performed correctly and decreases if it is incorrect

2.2. Participants

Due to the confidential nature of the manufacturing content, this study was conducted with employees from an aerospace company and a university within the United Kingdom. Twenty employees participated in the user study (20% female, mean age of 29.95 ± 12.87 SD years). Participants did not have extensive manufacturing knowledge prior to their participation, scoring a mean of 1.01 ± 1.2 SD in a scale from 0 to 3 (0-no experience, 1-minimal experience of less than one year, 2-moderate experience between one to three years, 3-extensive experience of over three years), and none previously trained on the assembly procedure studies in this article.

Each participant was asked to complete the VARK questionnaire (version 7.8) before participating. This questionnaire was designed by Fleming (2015) for the assessment of an individual’s learning

preference. According to VARK the main categories of learning preference are Visual, Aural, Read/Write, and Kinaesthetic but a user may be a multi-modal learner – when they have more than one main preference. In this study, participants were found to be in all of the main preference categories (Visual 20%, Aural 10%, Read/Write 10%, Kinaesthetic 45%), and two multi-modal categories (Visual-Kinaesthetic 10%, and Visual-Read/Write 5%). To minimise the influence of an individual’s learning preference in the experiment, participants of the different learning types were counterbalanced and randomly distributed through the two experimental conditions.

2.3. Procedure

In the experiment, participants underwent two sequential phases: (i) Study Phase and (ii) Evaluation Phase (Figure 8).

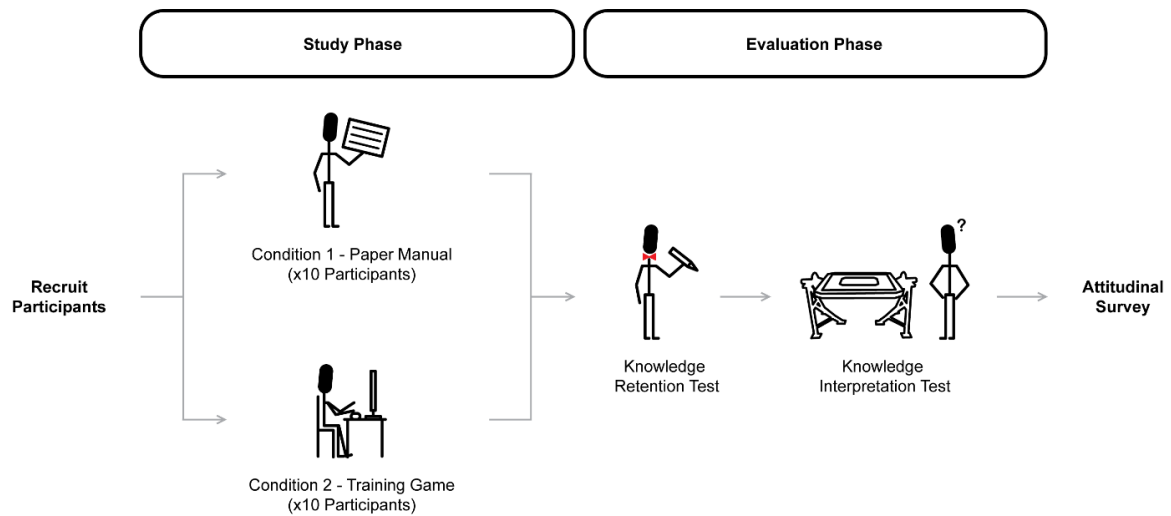


FIGURE 8. Participants were recruited for the experiment which consisted of two phases: 1) In the Study Phase half of the participants were asked to study the Paper Manual and the other half studied the Training Game, 2) in the Evaluation Phase all the participants took part in the Knowledge Retention Test and a Knowledge Interpretation Test, at the end of the experiment participants were asked to complete an attitudinal survey

2.3.1. Study Phase

In the Study Phase participants were given 30 minutes to study the manufacturing procedure. They were permitted to take notes during this time, but the notes could not be used in the evaluation phase. Participants were randomly assigned to one of two test groups:

- **Controlled Group:** Participants were trained using the PM (detailed in Section 2.1.1).
- **Experimental Group:** Participants were trained using the TG (detailed in Section 2.1.2). The TG was played on a desktop computer.

2.3.2. Evaluation Phase

The Study Phase was followed by an Evaluation Phase where participants were asked to complete two different knowledge tests to evaluate the learning effectiveness of the two instructional approaches. To ensure that the tests provide an accurate assessment of the desired learning outcome, the questions were designed with the help of a Subject Matter Expert in complex manufacturing. Participants were evaluated on two learning objectives:

1. **Knowledge Retention Test:** This was a written test in which participants completed a multiple choice test. The test is comprised of 14 questions, each correct answer corresponds

to one score. The highest score that can be achieved is 14. This test was designed to evaluate how much factual knowledge was retained from the Study Phase (Objective 1).

2. **Knowledge Interpretation Test:** This was a practical test that evaluated how participants applied the knowledge they had learnt in a real world environment (Objective 2). Participants were asked to perform two operations that included interacting with multiple moving parts on a scaled model (Figure 9). A score is awarded for each correct part located. A total of 33 scores can be achieved. This test was designed to 1) reflect real-life situations where parts could be missing or misplaced on equipment, and for 2) detecting whether a participant was applying acquired knowledge or simply relying on inference.



FIGURE 9. A movable cardboard model of the jig was created for the Knowledge Interpretation Test, the image on the left (Normal View) shows the jig in the standard position, the image on the right (Rotated View) shows the underside of the jig

2.3.3. Attitudinal Survey

At the end of the experiment participants were asked to complete an Attitudinal Survey. This survey was devised to evaluate the participants' subjective experience during three phases of the study (Study Phase, Evaluation Phase – Knowledge Retention Test, Evaluation Phase – Knowledge Interpretation Test). Table 2 presents the questions in the Attitudinal Survey. Participants answered a survey of 21 items, with 7 items referring to each phase of the experiment. Answers were provided in a seven-point Likert scale (1-Strongly disagree, 2-Disagree, 3-Somewhat disagree, 4-Neither agree nor disagree, 5-Somewhat agree, 6-Agree, 7-Strongly agree).

TABLE 2. Attitudinal Survey Questions

	Study Phase	Evaluation Phase: Knowledge Retention Test	Evaluation Phase: Knowledge Interpretation Test
Confidence	<ul style="list-style-type: none"> • I had enough time to study the assembly process. • There are technical terms that I did not understand. • At the end of the study phase I can fully recall the assembly process. • The work instructions were easy to understand. • The assembly process was not well explained. 	<ul style="list-style-type: none"> • I did well in the knowledge retention test. • The questions were relevant to the work instructions I studied. • The questions were difficult. • The work manual/video game provided enough information for me to answer the questions. • I guessed most of the answers. • The wording of the questions is too technical. 	<ul style="list-style-type: none"> • I made a guess on how the tasks are performed. • I did well in the knowledge interpretation test. • The questions were difficult. • The work manual/video game provided enough information for me to perform the tasks. • The wording of the questions is too technical. • The questions were relevant

	<ul style="list-style-type: none"> • I did not have time to go through the whole assembly process. 		to the work instructions I studied.
Engagement	<ul style="list-style-type: none"> • I enjoyed this part of the study. 	<ul style="list-style-type: none"> • I enjoyed this part of the study. 	<ul style="list-style-type: none"> • I enjoyed this part of the study.

The Attitudinal Survey investigates two elements: confidence level and engagement. The part of the survey that measures confidence level consisted of 50% positive and 50% negative statements. The statements representing a positive experience were scored positively (1 to 7), and the statements that represented a negative experience were scored negatively (-1 to -7). The average score indicated the participant’s overall confidence level. Regarding engagement, participants were asked to what extent they agreed with the following statement: I enjoyed this part of the study.

3. RESULTS

3.1. Knowledge Retention Test

The between-subjects one-way analysis of variance (ANOVA) of the Knowledge Retention score with the experimental conditions (PM and TG) shows no significant differences between the two conditions ($F(1,18) = 1.35, p = 0.266$) (Figure 10b).

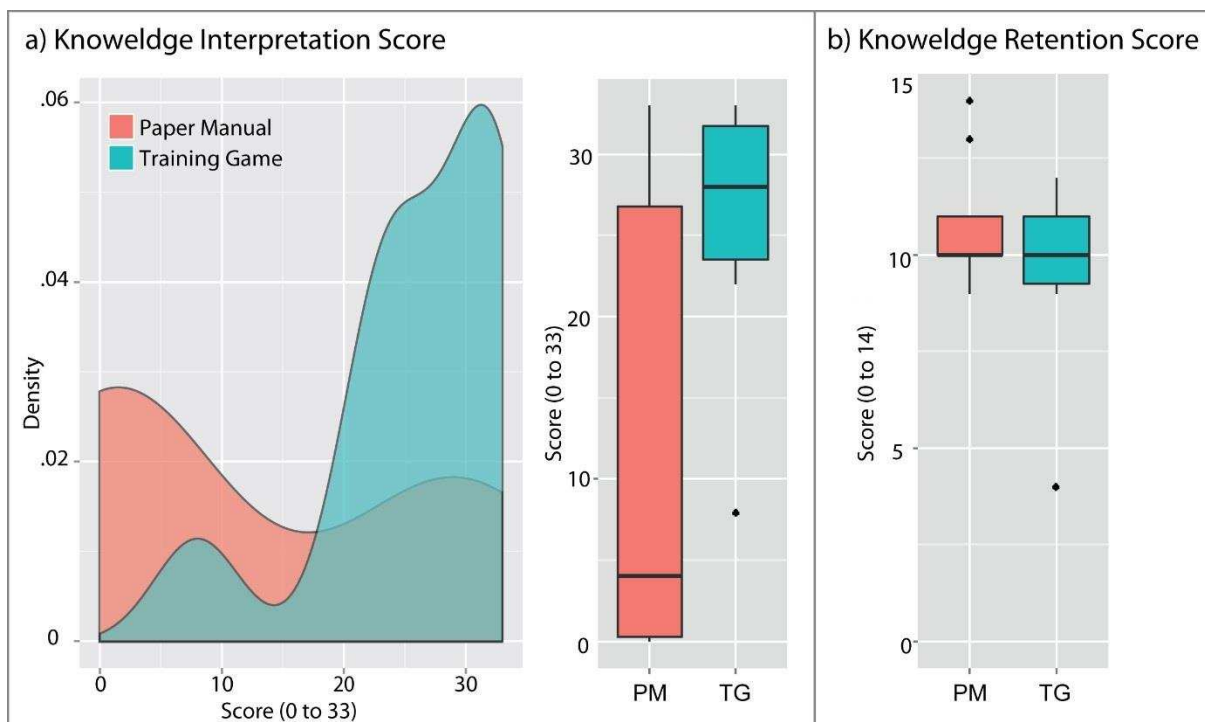


FIGURE 10. Graphs showing: a) Knowledge Interpretation score (0 to 33) for the two experimental conditions, boxplot and participants density map and; b) Knowledge Retention score (0 to 14) boxplot for the two conditions. The boxplots show the minimum, maximum, the quartiles and the median values (horizontal line) for each condition, the points represent the outliers.

3.2. Knowledge Interpretation Test

The between-subjects ANOVA with the Knowledge Interpretation score and the experimental conditions (PM and TG) shows a significant main effect ($F(1,18) = 6.933, p = 0.016$). Figure 10a shows the difference between the results of the PM condition ($M = 12.6, SD = 14.5$) and the TG

condition ($M = 26.3$, $SD = 7.7$), being significantly higher in the game condition. The density map (Figure 10a) shows the distribution of scores in each experimental condition. The peak of the density plot shows where there is the highest concentration of scores. For the TG condition, the peak is at 31, whereas the peak for the PM condition is at 2.

3.3. Demographic Factors

Demographic factors such as the manufacturing experience, gaming exposition, and the learning type (VARK) did not show significant interactions when added to the ANOVA model, indicating that these factors did not significantly influence the Knowledge Retention and Interpretation of the participants for either experimental condition.

3.4. Attitudinal Survey

3.4.1. Confidence

The average score on the Attitudinal Survey indicated the participant's overall confidence during the phases. Interestingly participants of the TG condition ($M = 3.53$, $SD = 1.6$) were significantly more confident during the Knowledge Interpretation Test than those in the PM condition ($M = 1.4$, $SD = 2.3$) (Wilcoxon rank sum test, $W = 23$, $p = 0.044$) (Figure 11a). No significant subjective differences were found for the Knowledge Retention Test or the Study phase.

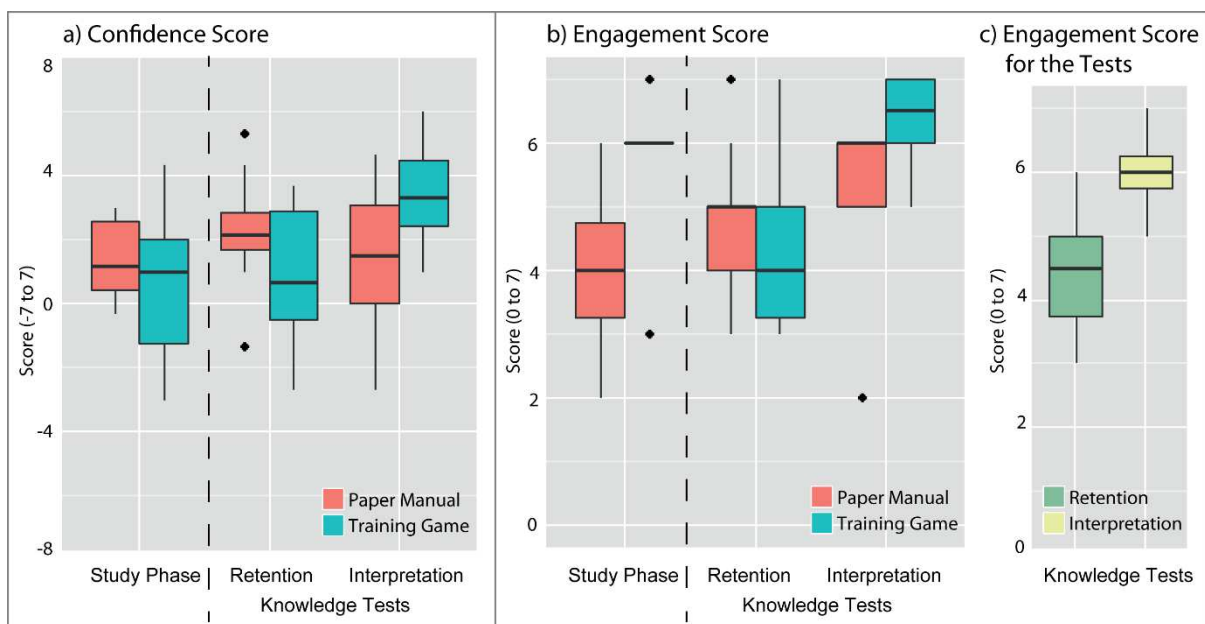


FIGURE 11. Attitudinal Survey results: a) boxplot diagram showing the confidence score (-7 to 7) for each phase and condition; b) boxplot diagram showing the engagement score (0 to 7) for each phase and condition; c) boxplot diagram showing the engagement score for the two tests. The boxplots show the minimum, maximum, the quartiles and the median values (horizontal line) for each condition, points represent the outliers.

3.4.2. Engagement

Regarding engagement in the Study Phase, participants in the TG condition ($M = 5.9$, $SD = 1.1$) scored significantly higher than those in the PM condition ($M = 4$, $SD = 1.4$) (Wilcoxon rank sum test $W = 14.5$, $p < 0.006$) (Figure 6b), i.e. engagement levels in the TG were higher than the PM. Furthermore, we find that participants overall enjoyed more the Knowledge Interpretation Test ($M = 5.7$, $SD = 1.4$) than the Knowledge Retention Test ($M = 4.5$, $SD = 1.2$) (Wilcoxon rank sum test, $W = 96$, $p = 0.004$) (Figure 11c).

Participants in the TG condition ($M = 6.4$, $SD = 0.7$) scored significantly higher in the engagement part of the survey than those in the PM condition ($M = 5$, $SD = 1.6$) for the Knowledge Interpretation Test (Wilcoxon rank sum test $W = 19$, $p = 0.012$) (Figure 6b), but not for the Knowledge Retention Test, where participants from both conditions (game: $M = 4.7$, $SD = 2.5$; paper: $M = 4.4$, $SD = 1.3$) had similar levels of engagement (Wilcoxon rank sum test $W = 58$, $p = 0.55$) (Figure 11b).

3.4.3. Participant Comments of the Two Conditions

Aside from the confidence and engagement levels, participants' opinions of the two study methods were collected in the Attitudinal Survey. They were asked to provide written feedback to an open-ended question: What did you like or dislike about the study session?

In the TG condition, participants responded positively to the gaming environment: *"I liked working in game environment"*, *"The experience was good"*. A number of suggestions were also offered to improve the game design. For example, the TG was designed to score the player immediately after a task meaning the player cannot replay a task unless they start the game again. One participant found *"the inability to go back to a previous incomplete process was a bit disappointing"*. The TG also seems to have a steep learning curve: *"Not familiar with gaming environment, does not know where to look for documents/instructions. Not used to this type of interface"*, leading to some participants proposing more didactic features: *"The game can give an overview of the process as a video in the first instance"* and *"More emphasis on key points could be made, not walls of text"*.

In the PM condition, participants found main hinderance of learning to be the ambiguity of the visual information: *"Pictures were not very clear which made it difficult when it came to the practical test"*, *"Picture quality could be better but that's usually the case for instruction manuals"*, *"It was difficult to visualise the assembly process"*. In particular, one participant stated that the PM need to be more precise when referring to the position of parts and materials *"It would be better if there were diagrams referring to the location of the bolts. This is not clear in the manual"*. The technicality of the PM was also a challenge *"The wording of the instructions was not suitable for purpose"*.

4. DISCUSSION

In this study a serious TG was compared with a traditional PM as means of training new workers for a complex manufacturing process in the aeronautic industry. The instructional approaches were assessed based on two knowledge tests that were taken immediately after the manufacturing process was studied. The results show that the TG is comparable to the PM for learning factual-based information (Objective 1), as evident in the Knowledge Retention Test (Figure 10). Similar results were established in studies comparing SG and non-gaming application (Papastergiou, 2009), and SG with computerised flash cards (Sward, Richardson, Kendrick, & Maloney, 2008). More optimistic results were reported in the use of SG for teaching adolescent cancer-related knowledge (Beale, Kato, Marin-Bowling, Guthrie, & Cole, 2007), whereby significant improvements in knowledge acquisition were identified.

With regard to learning to apply procedural knowledge (Objective 2), the TG approach proved to be significantly more effective than the PM. On average, participants in the TG condition scored more than double that of the participants in the PM condition in the Interpretation Test (Figure 10). This result is in line with a previous study whereby students were trained to assess and prioritise medical conditions in emergency situations (Knight et al., 2010) and for teaching procedural knowledge in healthcare (Torrente et al., 2014).

Both Objective 1 and 2 were mapped based on Bloom's Taxonomy (Anderson & Krathwohl, 2001; Buchanan et al., 2011; Krathwohl, 2002). It should be noted that being able to apply information

(Objective 2) requires higher order thinking skills compared to remembering information (Objective 1). As lower order skills are prerequisites to higher order skills (Pollock, Chandler, & Sweller, 2002), it is evident that the TG is far superior to the PM in the context of complex manufacturing training. Even though TG and PM are comparable in imparting a low order thinking skill, the TG outperformed the PM in conveying a higher order skill.

From the Attitudinal Survey it can be seen that participants in TG condition were more confident and engaged in the Knowledge Interpretation Test compared with participants in the PM condition. Although no significant difference of confidence level was identified in the Study Phase, participants in the TG condition also reported higher engagement level here. Overall, participants were more engaged in the Knowledge Interpretation Test compared with the Knowledge Retention Test.

4.1.1. Limitations and Strengths

This user study has some limitations. The sensitive nature of the learning content restricted the recruitment of participants. Running the same study with a bigger participant sample could unveil stronger evidence for the learning effectiveness of the two approaches (TG compared with PM). The Knowledge Interpretation Test was not based on a real assembly jig, but a model representation. This could have affected the results because even though the model representation can be used to measure how participants apply procedural knowledge to an extent, it does not truly reflect how participants would react on a real jig. However, the model representation had all the moving parts implemented and was a scaled reproduction of the CATIA-CAD model, the same model was used in the evaluation phase for both conditions. The engagement level may have been influenced by the novelty of the TG approach. Participants are likely to be familiar with paper-based learning; therefore learning through a game may temporarily appear more compelling. Nevertheless, the authors believe the effects were not just due to the novelty but rather to the exercise of learning by doing. Additionally, the PM used in this study was originally designed with the assumption that the subject had prior knowledge in the area of manufacturing (which the participants in this study did not). In this line, scores are expected to be higher in more experienced participants.

Whilst a number of limitations have been identified, this study also has a number of strengths. Throughout the development of the TG and in the planning of this study, a Subject Matter Expert with extensive experience in manufacturing was involved, which gives credence to the quality of both the game and the validation process. The TG was derived from a current PM used in industry. Both learning methods consist of the same written manufacturing instructions. In addition, a number of demographic factors were analysed to examine the extent of their influence on the test results, including: learning preferences, manufacturing experience, and gender. As none of these factors were found to have had significant influence on the results, it adds to the value of this study despite only 20 volunteers participated.

4.1.2. Future Work

This user study focused on short-term retention. It would be appropriate to study the long-term retention of the two conditions, as previous research has found that the use of SG applications had different results in long and short term retention (Bergeron, 2008; Chittaro & Buttussi, 2015; Jackson, Dempsey, & Mcnamara, 2011). On the one hand, Chittaro and Buttussi (2015) found the use of immersive SG applications to be superior in long term knowledge retention to educate players in aviation safety, more precisely they found that the fear-arousing and engaging element of the game leading to better long-term retention, while the short-term knowledge gain was comparable in the two conditions. Similar results were found by Bergeron (2008) where SG showed no decay in test scores after six weeks compared with the traditional learning method. However, on the other hand, Jackson

et al. (2011) reported that students performed worst in a SG, but improved over time to produce comparable results with the non-game condition.

This study only gathered results of how much a participant understands the manufacturing process using a written document (PM) or a game developed from this document (TG). It would be beneficial to identify the types and frequency of defects that are logged in the current manufacturing procedure. Comparing this data with types of errors made in this study would provide greater insight as to whether the TG does teach the correct procedure and if it is likely to prevent or reduce errors made on shop floors. Furthermore, future research could investigate whether involving the use of the TG in the training process would improve learning effective compared with traditional hands-on training. This would help to determine if the TG does in fact reduce the learning curve for workers going through the training process.

Another factor to consider in future research is the cost-effectiveness of creating a SG. Whilst cost-effectiveness has often been used as an argument for the creation of SG, perhaps due to the limited resources or difficulties in creating a real world environment, digitising training is an expensive process. To consider whether it is economically viable to create SG for complex manufacturing training, it would be beneficial to identify if the development cost can be justified by the prospective reduction in the COPQ, the reduction of cost of training new and existing workforce, and the reduction of cost through standardisation.

5. CONCLUSIONS

This user study shows evidence that the serious TG being evaluated is an effective method for training procedural knowledge in a complex manufacturing scenario. The results demonstrated that participants in the TG condition scored significantly higher in the Knowledge Interpretation Test (a test on the application of procedural knowledge) while maintaining similar levels in the Knowledge Retention Test (a test on the retention of factual knowledge) when compared to participants in the PM condition. Furthermore, participants enjoyed significantly more learning by playing than learning by reading, accordingly the results showed an increased in engagement levels in the TG condition. The strengths and limitations of this study are detailed in the previous section and should be carefully considered for future activities.

The success of the TG shows a real potential for adapting serious gaming for complex manufacturing training. Three research questions are suggested for future research, including: 1) investigating the effectiveness of the TG for long term retention, 2) comparing the results of this study with real shop-floor data to identify whether the manufacturing errors are likely to be prevented, and 3) evaluating if it is economically viable to widely deploy SG for complex manufacturing training scenarios.

REFERENCES

- Abate, A. F., Guida, M., Leoncini, P., Nappi, M., & Ricciardi, S. (2009). A Haptic-Based Approach to Virtual Training for Aerospace Industry. *Journal of Visual Languages and Computing*, 20(5), 318–325. <http://doi.org/10.1016/j.jvlc.2009.07.003>
- All, A., Castellar, E. P. N., & Van Looy, J. (2015). Towards a Conceptual Framework for Assessing the Effectiveness of Digital Game-Based Learning. *Computers & Education*, 88(1), 29–37. <http://doi.org/10.1016/j.compedu.2015.04.012>
- All, A., Nuñez Castellar, P. E., & Looy, J. Van. (2014). Measuring Effectiveness in Digital Game-Based Learning: A Methodological Review. *International Journal of Serious Games*, 1(2), 3–21. <http://doi.org/10.17083/ijsg.v1i2.18>
- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Aziz, E.-S. S., Chang, Y., Esche, S. K., & Chassapis, C. (2014). Virtual Mechanical Assembly Training Based on a 3D Game Engine. *Computer-Aided Design and Applications*, 12(2), 119–134. <http://doi.org/10.1080/16864360.2014.962424>
- Bal, M. (2012). Virtual Manufacturing Laboratory Experiences for Distance Learning Courses in Engineering Technology. In 119th ASEE Annual Conference and Exposition (pp. 1–12). San Antonio, TX, 10-13 June 2012. Retrieved from <https://peer.asee.org/virtual-manufacturing-laboratory-experiences-for-distance-learning-courses-in-engineering-technology>
- Bartoli, G., Del Bimbo, A., Faconti, M., Ferracani, A., Marini, V., Pezzatini, D., ... Zilleruelo, F. (2012). Emergency Medicine Training With Gesture Driven Interactive 3D Simulations. In 2012 ACM International Workshop on User Experience in E-Learning and Augmented Technologies in Education, UXeLATE 2012, Co-Located With ACM Multimedia 2012 (pp. 25–29). Nara, 2 November 2012. <http://doi.org/10.1145/2390895.2390903>
- Beale, I. L., Kato, P. M., Marin-Bowling, V. M., Guthrie, N., & Cole, S. W. (2007). Improvement in Cancer-Related Knowledge Following Use of a Psychoeducational Video Game for Adolescents and Young Adults With Cancer. *The Journal of Adolescent Health*, 41(3), 263–70. <http://doi.org/10.1016/j.jadohealth.2007.04.006>
- Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of Serious Games: An Overview. *Advances in Human-Computer Interaction*, 2013, 1–11. <http://doi.org/10.1155/2013/136864>
- Bergeron, B. P. (2008). Learning and Retention in Adaptive Serious Games. In *Medicine Meets Virtual Reality 16, MMVR 2008* (pp. 26–30). Long Beach, CA, 30 January - 1 February 2008. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/18391250>
- Borsci, S., Lawson, G., & Broome, S. (2015). Empirical Evidence, Evaluation Criteria and Challenges for the Effectiveness of Virtual and Mixed Reality Tools for Training Operators of Car Service Maintenance. *Computers in Industry*, 67, 17–26. <http://doi.org/10.1016/j.compind.2014.12.002>
- Buchanan, L., Wolanczyk, F., & Zinghini, F. (2011). Blending Bloom's Taxonomy and Serious Game Design. In *International Conference on Security and Management* (Vol. 7018, pp. 1–6). Las Vegas, NV, 18-21 January 2011. Retrieved from <http://securedesigns.com/blending-blooms-taxonomy-and-serious-game-design/>

- Chittaro, L., & Buttussi, F. (2015). Assessing Knowledge Retention of an Immersive Serious Game vs. a Traditional Education Method in Aviation Safety. *IEEE Transactions on Visualization and Computer Graphics*, 21(4), 529–538. <http://doi.org/10.1109/TVCG.2015.2391853>
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A Systematic Literature Review of Empirical Evidence on Computer Games and Serious Games. *Computers & Education*, 59(2), 661–686. <http://doi.org/10.1016/j.compedu.2012.03.004>
- Dickinson, J. K., Woodard, P., Canas, R., Ahamed, S., & Lockston, D. (2011). Game-Based Trench Safety Education: Development and Lessons Learned. *Journal of Information Technology in Construction*, 16, 118–132. Retrieved from <http://www.itcon.org/paper/2011/8>
- Ekanayake, H., Backlund, P., Ziemke, T., Ramberg, R., & Hewagamage, K. (2010). Game Interaction State Graphs for Evaluation of User Engagement in Explorative and Experience-Based Training Games. In *International Conference on Advances in ICT for Emerging Regions, ICTer 2010* (pp. 40–44). Colombo, Sri Lanka, 29 September - 1 October 2010. <http://doi.org/10.1109/ICTER.2010.5643272>
- Fleming, N. (2015). VARK: A Guide to Learning Styles. Retrieved February 15, 2016, from <http://vark-learn.com/>
- Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, Motivation, and Learning: A Research and Practice Model. *Simulation & Gaming*, 33(4), 441–467. <http://doi.org/10.1177/1046878102238607>
- Girard, C., Ecalle, J., & Magnan, A. (2013). Serious Games as New Educational Tools: How Effective Are They? A Meta-Analysis of Recent Studies. *Journal of Computer Assisted Learning*, 29(3), 207–219. <http://doi.org/10.1111/j.1365-2729.2012.00489.x>
- Gonzalez-Franco, M., Gilroy, S., & Moore, J. O. (2014). Empowering Patients to Perform Physical Therapy at Home. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014* (Vol. 2014, pp. 6308–6311). Chicago, IL, 26-30 August 2014. <http://doi.org/10.1109/EMBC.2014.6945071>
- González-Franco, M., Peck, T. C., Rodríguez-Fornells, A., & Slater, M. (2014). A Threat to a Virtual Hand Elicits Motor Cortex Activation. *Experimental Brain Research*, 232(3), 875–887. <http://doi.org/10.1007/s00221-013-3800-1>
- Gredler, M. E. (2004). Games and Simulations and Their Relationships to Learning. In D. H. Jonassen (Ed.), *Handbook of Research for Educational Communications and Technology* (2nd ed., pp. 571–582). Mahwah, NJ: Lawrence Erlbaum Associates.
- Greuter, S., Tepe, S., Peterson, J. F., Boukamp, F., D'Amazing, K., Quigley, K., ... Wakefield, R. (2012). Designing a Game for Occupational Health and Safety in the Construction Industry. In *8th Australasian Conference on Interactive Entertainment: Playing the System, IE 2012*. Auckland, 21-22 July 2012. <http://doi.org/10.1145/2336727.2336740>
- Jackson, G. T., Dempsey, K. B., & McNamara, D. S. (2011). Short and Long Term Benefits of Enjoyment and Learning Within a Serious Game. In *15th International Conference on Artificial Intelligence in Education, AIED 2011* (pp. 139–146). Auckland, 28 June - 1 July 2011. http://doi.org/10.1007/978-3-642-21869-9_20

- Kang, J., & Jain, N. (2011). Merit of Computer Game in Tacit Knowledge Acquisition and Retention. In 28th International Symposium on Automation and Robotics in Construction, ISARC 2011 (pp. 1091–1096). Seoul, 29 June 2011 - 2 July 2011. Retrieved from http://www.iaarc.org/publications/proceedings_of_the_28th_isarc/merit_of_computer_game_in_tacit_knowledge_acquisition_and_retention.html
- Knight, J. F., Carley, S., Tregunna, B., Jarvis, S., Smithies, R., de Freitas, S., ... Mackway-Jones, K. (2010). Serious Gaming Technology in Major Incident Triage Training: A Pragmatic Controlled Trial. *Resuscitation*, 81(9), 1175–9. <http://doi.org/10.1016/j.resuscitation.2010.03.042>
- Krathwohl, D. R. (2002). A Revision of Bloom' s Taxonomy: An Overview. *Theory into Practice*, 41(4), 212–218. http://doi.org/10.1207/s15430421tip4104_2
- Lewis, M. A., & Maylor, H. R. (2007). Game Playing and Operations Management Education. *International Journal of Production Economics*, 105(1), 134–149. <http://doi.org/10.1016/j.ijpe.2006.02.009>
- Ma, M., & Zheng, H. (2011). Virtual Reality and Serious Games in Healthcare. In S. Brahmam & L. C. Jain (Eds.), *Advanced Computational Intelligence Paradigms in Healthcare 6. Virtual Reality in Psychotherapy, Rehabilitation, and Assessment* (pp. 169–192). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-17824-5_9
- Mital, A., Pennathur, A., Huston, R. L., Thompson, D., Pittman, M., Markle, G., ... Sule, D. (1999). The Need for Worker Training in Advanced Manufacturing Technology (AMT) Environments: A White Paper. *International Journal of Industrial Ergonomics*, 24(2), 173–184. [http://doi.org/10.1016/S0169-8141\(98\)00024-9](http://doi.org/10.1016/S0169-8141(98)00024-9)
- Mitchell, A., & Savill-Smith, C. (2004). *The Use of Computer and Video Games for Learning: A Review of the Literature*. London: Learning and Skills Development Agency.
- Mitgutsch, K., & Alvarado, N. (2012). Purposeful by Design: A Serious Game Design Assessment Framework. In *Proceedings of the International Conference on the Foundations of Digital Games, FDG 2012* (pp. 121–128). Raleigh, NC, 30 May - 1 June 2012. <http://doi.org/10.1145/2282338.2282364>
- O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of Learning Outcomes: Evidence From the Computer Games Literature. *The Curriculum Journal*, 16(4), 455–474. <http://doi.org/10.1080/09585170500384529>
- Ong, S. K., & Mannan, M. A. (2004). Virtual Reality Simulations and Animations in a Web-Based Interactive Manufacturing Engineering Module. *Computers & Education*, 43(4), 361–382. <http://doi.org/10.1016/j.compedu.2003.12.001>
- Papastergiou, M. (2009). Digital Game-Based Learning in High School Computer Science Education: Impact on Educational Effectiveness and Student Motivation. *Computers & Education*, 52(1), 1–12. <http://doi.org/10.1016/j.compedu.2008.06.004>
- Peck, T. C., Seinfeld, S., Aglioti, S. M., & Slater, M. (2013). Putting Yourself in the Skin of a Black Avatar Reduces Implicit Racial Bias. *Consciousness and Cognition*, 22(3), 779–87. <http://doi.org/10.1016/j.concog.2013.04.016>
- Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating Complex Information. *Learning and Instruction*, 12(1), 61–86. [http://doi.org/10.1016/S0959-4752\(01\)00016-0](http://doi.org/10.1016/S0959-4752(01)00016-0) Get

- Sitzmann, T. (2011). A Meta-Analytic Examination of the Instructional Effectiveness of Computer-Based Simulation Games. *Personnel Psychology*, 64(2), 489–528. <http://doi.org/10.1111/j.1744-6570.2011.01190.x>
- Sun, S. H., & Tsai, L. Z. (2012). Development of Virtual Training Platform of Injection Molding Machine Based on VR Technology. *The International Journal of Advanced Manufacturing Technology*, 63(5-8), 609–620. <http://doi.org/10.1007/s00170-012-3938-1>
- Sward, K. A., Richardson, S., Kendrick, J., & Maloney, C. (2008). Use of a Web-Based Game to Teach Pediatric Content to Medical Students. *Ambulatory Pediatrics*, 8(6), 354–359. <http://doi.org/10.1016/j.ambp.2008.07.007>
- Torrente, J., Borro-Escribano, B., Freire, M., del Blanco, A., Marchiori, E. J., Martinez-Ortiz, I., ... Fernandez-Manjon, B. (2014). Development of Game-Like Simulations for Procedural Knowledge in Healthcare Education. *IEEE Transactions on Learning Technologies*, 7(1), 69–82. <http://doi.org/10.1109/TLT.2013.35>
- Vizendo. (2014). Homepage. Retrieved May 5, 2015, from <http://vizendo.se/>
- Williams-Bell, F. M., Murphy, B. M., Kapralos, B., Hogue, A., & Weckman, E. J. (2015). Using Serious Games and Virtual Simulation for Training in the Fire Service: A Review. *Fire Technology*, 51(3), 553–584. <http://doi.org/10.1007/s10694-014-0398-1>
- Wouters, P., Nimwegen, C. Van, Oostendorp, H. Van, Spek, E. D. Van Der, & Games, S. (2013). A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games. *Journal of Educational Psychology*, 105(2), 249–265. <http://doi.org/10.1037/a0031311>
- Wouters, P., van der Spek, E., & van Oostendorp, H. (2009). Current Practices in Serious Game Research: A Review From a Learning Outcomes Perspective. In T. Connolly, M. Stansfield, & L. Boyle (Eds.), *Games-Based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and Effective Practices* (pp. 232–250). Hershey, PA: Information Science Reference. <http://doi.org/10.4018/978-1-60566-360-9.ch014>
- Zyda, M. (2005). From Visual Simulation to Virtual Reality to Games. *Computer*, 38(9), 25–32. <http://doi.org/10.1109/MC.2005.297>