Computers & Education 56 (2011) 21–35

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Evaluation of a game to teach requirements collection and analysis in software engineering at tertiary education level

Thomas Hainey*, Thomas M. Connolly, Mark Stansfield, Elizabeth A. Boyle

University of the West of Scotland, Paisley, Scotland, United Kingdom

ARTICLE INFO

Article history:
Received 1 July 2010
Received in revised form
30 August 2010
Accepted 13 September 2010

Keywords:
Games-based learning
Software engineering
Requirements collection and analysis
Evaluation
Pedagogy

ABSTRACT

A highly important part of software engineering education is requirements collection and analysis which is one of the initial stages of the Database Application Lifecycle and arguably the most important stage of the Software Development Lifecycle. No other conceptual work is as difficult to rectify at a later stage or as damaging to the overall system if performed incorrectly. As software engineering is a field with a reputation for producing graduates who are inappropriately prepared for applying their skills in real life software engineering scenarios, it suggests that traditional educational techniques such as role-play, live-through case studies and paper-based case studies are insufficient preparation and that other approaches are required. To attempt to combat this problem we have developed a games-based learning application to teach requirements collection and analysis at tertiary education level as games-based learning is seen as a highly motivating, engaging form of media and is a rapidly expanding field. This paper will describe the evaluation of the requirements collection and analysis game particularly from a pedagogical perspective. The game will be compared to traditional methods of software engineering education using a pre-test/post-test, control group/experimental group design to assess if the game can act as a suitable supplement to traditional techniques and assess if it can potentially overcome shortcomings. The game will be evaluated in five separate experiments at tertiary education level.

1. Introduction

Computer games are regarded by some educationalists as highly engaging and it is hoped that by exploiting their highly compelling features that they can be used to help people learn effectively. Games-based learning (GBL) has been applied in a wide variety of different fields including medicine (Beale, Kato, Marin-Bowling, Guthrie, & Cole, 2007; Lennon, 2006; Roubidoux, 2005), knowledge management (Christoph, 2007; Chua, 2005), military training (Schneider, Carley, & Moon, 2005), science and mathematics (Habgood, 2007; Nelson, 2007; Squire, Barnett, Grant, & Higginbotham, 2004; Young & Upitis, 1999) promotion of language education (Johnson and Wu, 2008; Rankin, Gold, & Gooch, 2006), software engineering, computer science and information systems (Ford & Minsker, 2003; Jain and Boehm, 2006; Oh Navarro and Van der Hoek, 2005; Shaw & Dermoudy, 2005; Waraich, 2004; Zhu, Wang, & Tan, 2007). A recent study by Boyle et al. (submitted for publication) examines the literature on computer games and serious games, focusing on the potential positive impacts of gaming, particularly with respect to learning value and skill enhancement. Over 7392 papers were identified in the review of the research literature between 1996 and 2009, confirming the surge of interest in this area. However, only 127 papers were empirical and once a quality measure was applied, only 64 empirical papers remained. The literature review shows that playing computer games confers a range of perceptual, cognitive, behavioural and affective and motivational impacts and outcomes and the most frequently occurring outcomes and impacts were affective and motivational followed by knowledge acquisition/content understanding. This reflects the parallel interests in games as an entertainment medium but increasingly their use for learning. The authors felt that there was a dearth of evidence and believed that much more research was required to support the use of computer games in education.

Requirements collection and analysis in computing education is a fundamental process in various different learning modules across tertiary education. It is integral to the success of any software project whether it is a case study at university level or whether it is a small or large project in industry. No other conceptual work is as difficult to rectify at a later stage or as damaging to the overall system if performed

* Corresponding author. Tel.: +44 0 7814230679.
E-mail addresses: thomas.hainey@uws.ac.uk (T. Hainey), thomas.connolly@uws.ac.uk (T.M. Connolly), mark.stansfield@uws.ac.uk (M. Stansfield), liz.boyle@uws.ac.uk (E.A. Boyle).

0360-1315/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compedu.2010.09.008

incorrectly (Brooks, 1987). It is suggested that typical software engineering courses fail to teach students some of the skills they require as professional software developers as lectures and courseworks do not adequately prepare graduates about the software development process and that more novel approaches are required to address shortcomings of traditional approaches (Ludi and Collofello, 2001; Oh Navarro, 2006; Oh Navarro, Baker, and Van der Hoek, 2004).

Some problems associated with traditional approaches to teaching requirements collection and analysis could potentially be overcome with a GBL solution as GBL is perceived as a highly engaging form of supplementary learning by some educationalists. One of the main problems associated with the field of GBL is the distinct lack of empirical evidence supporting the approach. This paper will make a contribution to the empirical evidence in the field by conducting 5 evaluation experiments comparing a GBL approach to a role-playing and paper-based approach for teaching requirements collection and analysis. The evaluations will take place in Higher Education (HE) and Further Education (FE) to fully encompass tertiary education level. In the United Kingdom, Higher Education (HE) refers to university level and Further Education (FE) refers to college level. As well as presenting the results of the individual studies, this paper will present a comparative analysis between the evaluations at HE and FE level. In the following sections we will discuss previous work including: the problems teaching requirements collection and analysis, the advantages and disadvantages of traditional, how some GBL can overcome some of these shortcomings and then a new evaluation framework developed by the authors to guide the evaluation process. We then discuss the game to simulate the process of requirements collection and analysis. We then present the evaluation results including an analysis of five studies. The first three studies are at HE level and the fourth study is at FE level. The fifth study compares the results of the HE and FE studies. Finally there will be a discussion of the results and future research directions.

2. Problems teaching requirements collection and analysis

We have discussed the issues associated with teaching requirements collection and analysis in Connolly, Stansfield, and Hainey (2007). In this section, we compare the problems associated with different approaches to teaching this type of subject matter. We then discuss how GBL can potentially overcome some of the shortcomings of traditional approaches.

2.1. Advantages and disadvantages of traditional approaches

Shaw and Dermoudy (2005) believe "students have little empathy for, or affinity with the fundamentals of software engineering practice when it is first introduced." Briggs (1994) highlights that students must be motivated to engage in the construction of software otherwise they will not use the taught techniques effectively when entering employment as it will be difficult to convince them of the effectiveness of the methods and techniques taught. Dawson (2000) emphasises that educational environments have experiential limitations, as they do not encompass the full range of experiences faced in reality. Some of the advantages and disadvantages of traditional approaches displayed in Table 1 are adapted from: Bonwell (1996), Cashin (1985), Wehrli and Nyquist (2003), Davis (2001), the ADPRIMA Instructional Methods information website (2009), Shaw and Dermoudy (2005) and Connolly, Stansfield, McLellan, Ramsay, and Sutherland (2004). The main traditional approaches addressed are lectures, role-play and paper-based case studies.

2.2. Overcoming some of the problems of traditional approaches

GBL can potentially help with some of the shortcomings as games enable meaning to be *situated* (Lave & Wenger, 1991), *anchored* (Bransford, Sherwood, Hasselbring, Kinzer, & Williams, 1990) and support "conceptual interaction" (Laurillard, 1996). Situated learning is important as GBL is at its most powerful when it is "personally meaningful, experimental, social, and epistemological all at the same time" (Shaffer, Squire, Halverson, & Gee, 2004, p. 3). Games provide experiential learning where knowledge is created through the transformation of experience (Kolb, 1984) providing the four stages of the experiential learning cycle: concrete experience, reflective observation, abstract conceptualisation, and active experimentation. Healy and Connolly (2007) compared the difference between traditional methods and GBL. Some of the main differences concern the passivity of traditional methods as opposed to the interactive and active learning strategies evident in GBL. Traditional methods are viewed as outdated and based on an instructivist methodology. GBL, on the other hand, is considered more up-to-date and underpinned by students actively seeking information and knowledge. A further and significant difference is that while traditional methods are proven, GBL is still relatively unproven (de Freitas, 2006; Connolly et al., 2007).

In terms of specifically teaching requirements collection and analysis, GBL can provide a constructivist learning environment where learners can practice the formulation of a requirements specification through requirements elicitation and learning by doing (Aldrich, 2005). The learner can be immersed in a simulated, practically-oriented, thickly authentic office environment utilising a software engineering epistemic frame. The environment will support communication structures by means of an email system, reflection and the formation of metacognitive strategies by appropriate feedback and allowing the learner to exercise eclectic decision making. The GBL environment will use Sommerville's (2006) definitions of user and system requirements and will automatically group the requirements into the proper category once collected. This will provide the learners with an appreciation of the differences of the two main different types of software requirement. The environment will have ill-defined, ambiguous and changing requirements allowing the learner the opportunity to generate additional questions and clarify requirements with the different characters in the game. The learner will be allowed to explore the environment until they are satisfied that they have collected all of the relevant requirements. The requirements specification will be refinable in the sense that redundant requirements can be deleted. The learner will be required to use their experience in a creative fashion in a holistic, people oriented environment to collect requirements in a non-linear fashion. The environment will contain multiple stakeholders that can be questioned and will also have interactive objects to attempt to enhance the learning experience.

Table 1
Advantages and disadvantages of traditional approaches to teaching requirements collection and analysis.

Teaching Technique	Advantages	Disadvantages
Lectures	• Lecturers have full control of the learning experience. • Intrinsic interest of a subject can be communicated through the lecturer's enthusiasm. • Large amounts of information can be tailored and presented to large audiences containing material not otherwise available to students. • Lectures can provide a model of how professionals address problems and questions. • Appeal to students who learn by listening and present very little risk for students.	• Lecturer is required to be an effective speaker and become a 'sage on stage'. • No mechanism exists to ensure that students are intellectually engaged, meaning that they are often passive recipients resulting in information being quickly forgotten. • Student's attention begins to wane after a short period (approximately 15–25 min). • Lectures are not suited for teaching abstract, complex subjects, or higher order thinking skills such as values, motor skills, analysis and application etc. • Assumes that all students are at the same level of understanding and learn at the same pace. • Students who have different optimal learning styles other than listening are at a disadvantage.
Role-Play	• Participants are actively involved in the exercise. • Enhances the learning experience by adding elements of variety, reality and specificity. • Provides a safe environment to increase practice experience when real life experiences are unavailable. • Immediate feedback is provided and can give new perspectives on situations. • Likelihood of transfer from the classroom to the real world is improved.	• Puts pressure on learner to perform and can result in embarrassment if learners are too self conscious. • Can be time consuming as good practice usual requires a debriefing session. • The role-play has to be well planned, monitored and orchestrated otherwise it may lack focus. • Puts mental pressure on the acting participants if questions are asked that deviate from the script. • Requires appropriate monitoring by a knowledgeable person to provide appropriate feedback.
Paper-based Case-Studies	• Students can apply new skills. • Develops analytical problem solving skills. • Can be relatively self-pacing if it is in the form of a coursework as participants can study in their own time. • Useful method for finding a large number of solutions for complex issues.	• The case study must be carefully prepared and defined. • It may be difficult for the students to see the relevance of the case study for their own situation. • Inappropriate results may occur if the case study has insufficient information. • Approach is not suitable to elementary education.

3. Simulating requirements collection and analysis using a games-based learning approach

3.1. Game play

The basic idea of the game is for the team (comprising one or more players) to manage and deliver a number of software development projects. Each player has a specific role, such as project manager, systems analyst, systems designer or team leader. A bank of scenarios have been created based on case studies the authors have been using for many years in teaching and learning; for example, the Dream Home Estate Agency, the StayHome Online DVD Rentals company and the Perfect Pets Veterinary Clinic, the Blackwood Library and the Fair Winds Marina. Each scenario has an underlying business model; for example, there will be a budget for the delivery of the project, a set timescale for the delivery of the project and a set of resources (for example, staff with specified technical specialisations) that can be used on the-project. Additional resources can be brought in for a project although this will have a cost and timescale (delay) associated with it. The project manager has overall responsibility for the delivery of each project on budget and on time and is given a short brief for each project. Communication is one of the key aspects of the game and the project manager must communicate relevant details of the project to the other players. This will be done using a message metaphor – any player can communicate with any other player(s) by sending a message. Players have a message board that indicates whether there are any unread messages.

The player(s) assigned to the system analyst role has to identify the requirements for the project. To do this, the player must move through the game and 'talk' to the non-player characters (NPCs) in the game, as illustrated in Fig. 1. In addition, there are objects in the game that can also convey relevant information when found (for example, a filing cabinet may convey requirements). For the prototype game we are using written transcripts in place of NPC speech. We hope shortly to use lipsynching within the game to have the NPCs 'talk' to the system analyst. Each NPC's 'speech' will contain some general background details and a number of requirements (the analyst has to distinguish the requirements from the general details). Visiting the same NPC may generate the same speech or a new speech. Each speech will generate a transcript that the analyst can visit at any point in the game. The transcript is presented as a numbered list of requirements. During the play, the analyst can use the transcripts to produce an initial 'wishlist' of requirements, which can be refined until such time as the analyst believes all requirements have been identified, at which point the analyst can send the completed requirements to the project manager. The project manager now has two choices: send the requirements to the designer to produce an outline high-level design or consider the requirements to be incorrect and ask the analyst to rework the requirements (asking for rework will have a 'cost' associated with it).

4. Evaluation of the requirements collection and analysis game

All evaluations were conducted using the evaluation framework proposed by Connolly, Stansfield, and Hainey (2009). For the purposes of this study the reported results will particularly focus on pedagogy, aspects of the approaches (game, role-play and paper-based) and perceptions.

Fig. 1. Screen during requirements collection.

4.1. Methodology

The methodology selected to evaluate the RCAG is a pre-test → post-test, experimental/control group design where the control group was a traditional teaching approach. Fig. 2 shows the experimental design selected. For the randomisation methodology, the HE students were randomly assigned to the experimental or control group. The FE students were randomly assigned to an experimental or control group based on institution. This was primarily due to time constraints as there was only sufficient time allocated in the FE institutions to do one part of the evaluation i.e. role-play, game or paper-based.

The measurements collected for each group were: results to knowledge questions for the pre and post-tests for learning effectiveness, ranked ordinal data on a scale of 1–5 (5 being the highest and 1 being the lowest) for aspects ratings and perceptions. There was a slight difference in the aspect ratings collected between the groups as some of the aspects were not comparable, for example, a rating for graphics could be collected in the game group but could not be collected in the role-play or paper-based groups.

The results obtained were analysed using SPSS version 15. Non-parametric statistical tests were used to analyse the data. Wilcoxon matched-pairs signed, ranks tests were used to assess if there was an increase in knowledge within the individual groups. Kruskal–Wallis tests were used to assess if there was any significant difference in learning between the individual groups. Descriptive statistics were used to rank the aspects of the particular groups i.e. game group, role-play group and paper-based group. Mann–Whitney *U* tests were used to assess differences in levels of knowledge, rankings of aspects and perceptions between FE and HE groups.

4.2. Participants

The combined experiment involved 92 students, 55 students from HE and 37 students from FE. Table 2 shows the list of participants. The majority of the participants had a small, a very small or no instruction in requirements collection and analysis.

4.3. Procedure

4.3.1. Experimental group (game group)

In the experimental game group, each participant was presented with a pre-test designed to collect some demographic and learner type information, assess the level of knowledge the participant already possessed about requirements collection and analysis and collect

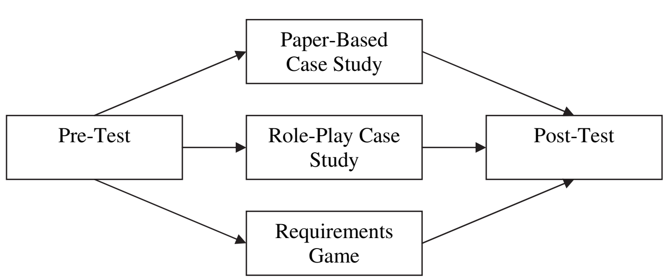

Fig. 2. Selected experimental methodology.

Table 2
List of participants from Higher Education and Further Education.

Education Level	Institution and Course	Group	Details
HE Participants		$n = 55$	
Study 1	University A Introduction to Database Systems Module	Experimental, $n = 16$	14 participants were male, 2 were female. Mean age was 25.13 years (SD = 5.57) with a range from 19 to 36.
		Control (role-play group), $n = 13$	8 (61.5%) postgraduate and 5 (38.5%) undergraduate. Mean age of the participants was 29.31 (SD = 7.96) with a range from 19 to 44 years.
	University A: PhD students from various disciplines	Control (paper group), $n = 10$	8 (80%) were female and 2 (20%) were male. Mean age was 30.90 (SD = 6.24) with a range from 24 to 44 years.
Study 2	University A: Computer Games Technology Module	Experimental, $n = 4$ Control, $n = 4$	All participants were male. Mean age was 21.50 (SD = 0.76) with a range from 21 to 23.
Study 3	University B: Computer Games Technology Module	Experimental, $n = 4$ Control, $n = 4$	All participants were male. Mean age was 19.13 (SD = 1.89) with a range from 17 to 22.
FE Participants		$n = 37$	
Study 4	College A: HNC Computer Games Design	Experimental, $n = 7$	34 participants were male. The mean age was 22.61 (SD = 6.75) with a range of 18–47.
	College B: HNC Multimedia	Experimental, $n = 7$	
	College C: HND Technical Support	Experimental, $n = 9$	
	College D: HND Software Development	Control, $n = 14$	
HE/FE participants		$n = 82$ (45 from HE, 37 from FE)	
Study 5 Comparison of HE and FE	University A, B and College: A, B, C, D.	Experimental, $n = 47$ Control, $n = 35$	24 from HE and 23 from FE 21 from HE and 14 from FE

additional information, such as the most important aspects that they believe a game should possess. The participant was then presented with a summary information instruction screen that could be referred to at any point during game play. The instructions screen detailed how to operate within the game environment and described the different types of requirements. The participant played the game for as long as it took to produce a requirements specification that the player believed to be satisfactory. This generally took between 15 and 40 min. Upon completion of the game the participant completed a post-test designed to test knowledge of requirements collection and analysis to assess whether there was an increase in knowledge. The knowledge questions were not identical to the pre-test but very similar to ensure that the participants did not learn from the previously asked questions present in the pre-test. It also collected information about positive and negative aspects of the game, aspect ratings, perceptions and preferences of the learners.

4.3.2. *Control groups (role-play and paper-based case study)*

Study 1 used two control groups, one role-play and the other paper-based for the purpose of identifying the best traditional teaching approach to compare with the game. Due to the interactive and textual nature of the RCAG, the two most similar approaches identified were role-play and paper-based. Lectures and live-through case studies were not considered similar enough as lectures were not sufficiently interactive and live-through case studies were too time-consuming.

In the control role-play group, each participant was presented with a pre-test similar to those in the game group collecting demographic information and learner type information and assessing the level of knowledge of requirements collection and analysis that the participants possessed. The knowledge test was identical to the experimental game group. The tests were slightly different because of having to adapt the tests for the particular contexts (i.e. game, role-play and paper-based) as some aspects are not comparable, for example graphics can be evaluated in a computer game, but can not be evaluated in a paper-based study or a role-play study. The participants were then provided with a summary for the role-playing exercise similar to the summary screen for the game but with all the game elements removed in the paper-based format. As in the game exercise, participants were provided with questions to start the role-play. Members of staff and students acted out the parts of the characters in the game using the exact same case study and script that was incorporated into the game. Participants were told to note down any requirements and any additional questions they asked that were not included in the script. The role-play exercises took approximately 15 min–90 min and proceeded until all characters were interviewed to the satisfaction of the participants. Participants then completed the post-test to assess whether there was an identifiable increase in knowledge. The questions were not identical with those in the pre-test to prevent the participants having advanced knowledge of the questions, however the questions were similar. The knowledge test questions in the role-play post-test were identical to those in the experimental group post-test. It also collected information about the positive and negative aspects of the activity, aspect ratings, perceptions and preferences.

In the control paper-based group each participant was presented with the pre-test collecting demographic information, learner type information and assessing the level of knowledge similar to the other two groups. The participants were then presented with the same summary information excluding game and role-play elements. The script for the role-play and the game were provided to the participants and they were asked to underline all requirements. They were also asked to note down any additional questions that they believed to be necessary. Participants were then presented with the post-test similar to those used in the other two groups but adapted for a paper-based study. The knowledge test questions were the same as in the other two groups.

4.4. *Study one in HE (University A)*

4.4.1. *Learning effectiveness*

Mann–Whitney *U* tests indicated that there was no significant difference in knowledge levels in the pre-tests between the three groups. The mean score on the pre-test for the game group was 3.86 out of 6 (SD = 1.36) and on the post-test was 5.00 out of 6 (SD = 1.26).

A Wilcoxon matched-pairs signed, ranks test showed that the scores following the game intervention were significantly higher than scores before the game intervention ($Z = -2.675$, $p < 0.001$). For the role-playing group the pre-test mean on the knowledge score was 3.69 out of 6 (SD = 1.03) and on the post-test 4.62 out of 6 (SD = 1.39). A Wilcoxon matched-pairs signed, ranks test showed that the scores following the role-play intervention were also significantly higher than scores before the intervention ($Z = -2.053$, $p < 0.05$). The mean score for the pre-test in the paper-based case study group was 3.50 out of 6 (SD = 1.72) and 4.50 out of 6 (SD = 1.08) on the post-test. A Wilcoxon matched-pairs signed, ranks test showed that the scores following the paper-based intervention were not significantly higher than scores before the intervention ($Z = -1.549$, $p > 0.05$). A Kruskal–Wallis test between the difference in the knowledge scores of the pre and post-test was not significant ($\chi^2 = 0.272$, $p > 0.05$). This indicates that the variation of score-difference within groups is larger than the differences between the groups indicating that while there was an increase in knowledge in the role-play and game groups, no group significantly outperformed the other.

4.4.2. Aspect ratings experimental group (game) and control groups

Participants' mean ratings of the aspects of the game-based learning environment (from very good, scored as 5 to very bad, scored as 1) were generally positive suggesting that participants were generally satisfied with the game interface as very few marks of 2 or less were observed. Participants' ratings of the aspects of the role-play activity (from very good scored as 5 to very bad scored as 1) were also generally positive. Participants' ratings of the aspects of the paper-based case study (from very good, scored as 5 to very bad, scored as 1) were not particularly positive. The realism of narrative, dialog and characters being were rated highly which is encouraging as the paper-based script is a direct representation of the dialog in the game, indicating that the case study is acceptable regardless of what format it is delivered in. Table 3 shows the comparison of the different aspects between the three groups.

4.4.3. Learner perceptions

4.4.3.1. Experimental group (game). The students were generally supportive of the use of the game in learning. 15 of the 16 participants agreed or strongly agreed that the game could be used in the intended domain in the future and 14/16 felt that the game fitted well into its intended context. 12 agreed that the game was engaging, 9 agreed that the game sustained their engagement, 9 would like to play the game again and 12 believed that the game could increase their skills at HE level. Students were asked the open-ended question of what they thought were the most positive aspects of the game. Several students picked up on the graphics as a positive feature of the game, while several liked the interaction and the conversational nature of the requirements collection:

- "The interaction.", "Conversation, collection of data.", "It shows the user how and from where s/he could collect the required data.", "Different people to ask e.g. advisor, manager."

One participant identified the real world setting of the game as useful in making links between being a student and working:

- "It is good that it feels like you work in a company and have to solve a 'real' job."

One student commented on the game as useful in learning:

- "Makes you think so helps you learn."

Participants also offered useful feedback about how to improve the game:

- "Speed of character, how all the questions piled up/difficulty of organisation. Information overload at the beginning/no back story.", "It was too repetitive.", "No real control of the characters.", "I couldn't find the right persons for solving the conflicts I found.", "I think it must have more controls and more questions to ask."

4.4.3.2. Control group (role-play). The students were generally supportive of the role-playing exercise in learning as well. 9 of the 13 who participated agreed or strongly agreed that the role-play exercise could be used in the intended domain in the future and 7/13 felt that the role-play fitted well into its intended context. 6 agreed that the role-playing activity was engaging, 5 agreed that the activity sustained their engagement. 5 would like to participate in a similar activity and 8 believed it could increase their skills at HE level.

Table 3
Comparison of aspects between the groups.

	Experimental Group (Game)	Control Group (Role-Play)	Control Group (Paper-Based Case Study)
Realism of the environment/realism of the scenario	4.00, SD = 0.82	3.31, SD = 0.95	3.70, SD = 1.41
Graphics	3.75, SD = 0.86	*****	*****
Navigation inside the environment	3.75, SD = 0.93	*****	*****
Character customisation	3.38, SD = 0.89	*****	*****
Realism of the characters in the environment/characters and the dialog	3.31, SD = 1.01	3.00, SD = 1.15	3.40, SD = 1.51
Clear goal structure	3.31, SD = 1.08	3.54, SD = 1.13	2.70, SD = 1.83
Narrative and dialog	3.25, SD = 1.06	3.54, SD = 0.78	4.20, SD = 0.63
Help and scaffolding facilities	3.13, SD = 0.89	2.85, SD = 1.14	2.20, SD = 1.75
Collaboration	2.37, SD = 1.75	3.77, SD = 0.832	2.00, SD = 1.70
Sound	1.19, SD = 1.47	*****	*****
Ability to improvise	*****	3.46, SD = 0.88	2.00, SD = 1.63

Students were asked the open ended question of what they thought were the most positive aspects of the role-play activity. Several students picked up on the interactive nature of the role-play activity:

- "Interactivity.", "Interacting with others.", "Group interaction."

3 students commented on the collecting information aspect of the exercise:

- "Collecting important info.", "Getting a lot of information quickly.", "Questions and answers."

2 students picked up on the hierarchical nature of the case study:

- "See how different employees at different levels in the company might operate.", "Finding out about the database, find out different views."

Students also commented on the negative aspects of the role-playing exercise. Two students picked up on the limitations of having a script with a real world interaction:

- "Not all of the information asked for was contained in the script.", "Some questions weren't relevant to some members of staff."

Two students picked up that it is difficult to record information when people are talking at speed:

- "Pushed for time writing down when people speak.", "Not getting enough information."

Some students found the experience confusing:

- "Not sure what was required.", "Can get quite confounded."

4.4.3.3. Control group (paper-based case study). The students were not as supportive of the paper-based case study exercise in learning. 5 of the 10 who participated agreed that the paper-based exercise could be used in the intended domain in the future and 4/10 felt that the paper-based exercise fitted well into its intended context. 5 agreed that the paper-based activity was engaging, 5 agreed that the activity sustained their engagement. 4 would like to participate in a similar activity and 6 believed that it could be used to increase skills at HE level. Students were asked the open ended question of what they thought were the most positive aspects of the paper-based case study activity. 2 students commented on the different views in the case study:

- "The breakdown of the scripts in term of users.", "Seeing the different requirements/opinions of the various users about what they felt the database should provide." One student commented on being able to re-read the case study: ·"Being able to re-read the case study as I worked through it."

One student commented on the pace of the exercise:

- "Dictating my own pace."

Students also commented on the negative aspects of the paper-based case study exercise. 4 students commented on the length of the exercise:

- "Length of the exercise.", "Long and confusing.", "Length of it.", "Quite long." One student indicated that the exercise did not hold his or her attention: "Doesn't interest me or hold my attention."

4.5. Study two in HE (University A)

4.5.1. Learning effectiveness

A Mann–Whitney *U* test indicated that there was no significant difference in knowledge levels in the pre-test between the game and role-play groups. Participants' scores on the knowledge requirements tests before and after both interventions were compared. The mean score on the pre-test for the game group was 7.50 out of 13 (SD = 4.72) and on the post-test 11.25 out of 13 (SD = 1.71). A Wilcoxon matched-pairs signed, ranks test showed that the scores following the game intervention were significantly higher than scores before the game intervention ($Z = -1.604$, $p < 0.05$). The mean score in the pre-test for the role-play was 6.25 out of 13 (SD = 1.5) and on the post-test 8.50 out of 13 (SD = 1.00). A Wilcoxon matched-pairs signed, ranks test showed that the scores following the role-play intervention were significantly higher than scores before the role-play intervention ($Z = -1.841$, $p < 0.05$). A Kruskal–Wallis test between the difference in the knowledge scores of the pre and post-test was not significant ($\chi^2 = 0.021$, $p > 0.05$). This indicates that the variation of score-difference within groups is larger than the differences between the groups. This means that while both groups increased their knowledge, no one group significantly outperformed the other.

4.5.2. Aspects ratings

4.5.2.1. Experimental group (game). The ratings of the aspects of the game were generally very positive, however it should be noted that the results in this case were collected from a small sample. Table 4 shows the results ranked from highest (5) to lowest (1).

Table 4
Ratings of the aspects of the game.

Aspect	Rank	Mean	SD
Navigation inside the environment	1st	4.25	0.5
Graphics	2nd	4.00	0.82
Realism of characters in the environment	2nd	4.00	0.0
Clear goal structure	2nd	4.00	0.82
Narrative and dialog	3rd	3.75	0.5
Character customisation	3rd	3.75	0.5
Realism of environment	3rd	3.75	0.5
Control mechanism and interface	3rd	3.75	0.96
Help and scaffolding	4th	3.50	0.58
Collaboration	5th	1.00	1.41
Sound	6th	0.25	0.5

4.5.2.2. Control group (role-play). The ratings of the aspects of the role-play were not particularly positive in comparison with the game, however it should be noted that the results were collected from a small sample. Table 5 shows the results ranked from highest (5) to lowest (1).

4.5.3. Learner perceptions

4.5.3.1. Experimental group (game). The students were generally supportive of the game in learning. 3 of the 4 participants agreed that they would be willing to play the game for a prolonged period of time. 3/4 agreed that they would play it as part of a computing course. 3 agreed that the game was engaging, however only 1 reported that it sustained their engagement. All participants agreed that the game fitted well into its intended context. 1 strongly agreed and 2 agreed that the game could be used in its anticipated domain in the future. All participants believed that future development of the game could potentially help learners of their demographic and 3 believed that future development of the game could potentially help them. All participants agreed that the game provided a good overview of requirements collection and analysis and all participants either agreed or strongly agreed that they were proficient at playing games. 3 agreed and 1 strongly agreed that the game could be used to increase skills to HE level. In terms of realism 2 agreed that the realism of the game was adequate.

4.5.3.2. Control group (role-play). The students were not very supportive of the role-play activity in learning. 2 of the 4 participants disagreed and 1 strongly disagreed that they would be willing to participate in the activity over a prolonged period of time. 3 strongly disagreed that they would be willing to participate in a role-play exercise as part of a computing course. Only 1 participant believed that the activity was engaging and only 1 participant believed that the activity sustained their engagement. All participants either disagreed or strongly disagreed that they would like to participate in a similar activity again. All participants were indifferent about the role-play fitting well into its intended context and 3 disagreed that the role-play exercise could be used in its anticipated domain in the future. 1 participant believed that future development of the role-play could potentially help learners of their demographic while 3 disagreed. All participants believed that the realism of the role-play was inadequate and that the role-play could not increase their skills. 3 participants disagreed that the role-play could be used in HE.

4.6. Study three in HE (University B)

4.6.1. Learning effectiveness

A Mann–Whitney *U* test indicated that there was no significant difference in knowledge levels in the pre-test between the game and role-play groups. Participants' scores on the knowledge requirements tests before and after both interventions were compared. The mean score on the pre-test for the game group was 4.50 out of 13 (SD = 3.11) and on the post-test was 9.75 out of 13 (SD = 1.26). A Wilcoxon matched-pairs signed ranks test indicated that the increase in knowledge scores on the pre-test and post-tests was significant ($Z = -1.604$, $p < 0.05$). The mean score in the pre-test for the role-play group was 6.75 (SD = 0.96) and on the post-test was 9.75 (SD = 0.96). A Wilcoxon matched-pairs signed ranks test indicated that the increase in knowledge scores on the pre-test and post-tests was significant ($Z = -1.841$, $p < 0.05$). A Kruskal–Wallis test between the difference in the knowledge scores of the pre and post-test was not significant ($\chi^2 = 0.788$, $p > 0.05$). This indicates that the variation of score-difference within groups is larger than the differences between the groups. This means that while both groups increased their knowledge, no one group significantly outperformed the other.

As the results of study 2 and 3 being relatively small samples, when the results of the two studies are combined i.e. 8 in the game group and 8 in the role-play group, the results are as follows: A Wilcoxon matched-pairs signed ranks test indicated that the knowledge increase in the game group was significant ($Z = -2.201$, $p < 0.01$). A Wilcoxon matched-pairs signed ranks test indicated that the knowledge increase in the role-play group was also significant ($Z = -2.539$, $p < 0.01$). A Kruskal–Wallis test between the difference in the knowledge scores of the pre and post-test was not significant ($\chi^2 = 0.635$, $p > 0.05$). This indicates that the variation of score-difference within groups is larger than

Table 5
Ratings of the aspects of the role-play.

Aspect	Rank	Mean	SD
Clear goal structure	1st	3.25	0.98
Realism of scenario	2nd	3	0.82
Narrative and dialog	3rd	2.75	0.96
Ability to improvise	3rd	2.75	0.5
Realism of characters in the dialog	4th	2.25	0.5
Help and scaffolding	5th	1.75	0.96
Collaboration	6th	1.75	0.96

the differences between the groups. This means that while both groups increased their knowledge, no one group significantly outperformed the other.

4.6.2. Aspects ratings

4.6.2.1. Experimental group (game). The ratings of the aspects of the game were generally positive, however it should be noted that the results in this case were collected from a small sample. Table 6 shows the results ranked from highest (5) to lowest (1).

4.6.2.2. Control group (role-play). The ratings of the aspects of the role-play were generally positive in comparison with study 2 however it should be noted that the results in this case were collected from a small sample. Table 7 shows the results ranked from highest (5) to lowest (1).

4.6.3. Learner perceptions

4.6.3.1. Experimental group (game). The students were generally supportive of the game in learning. 2 of the 4 participants agreed that they would be willing to play the game over a prolonged period of time and all participants agreed that they would be willing to play the game as part of a computing course. 3/4 of the students agreed that the game was engaging however only 2 agreed or strongly agreed that the game sustained their engagement. 2 participants strongly agreed and 2 agreed that the game fitted well into its intended context and all participants believed that the game could be used in its anticipated domain in the future. All participants believed that future development of the game could potentially help learners of their demographic and 3 participants believed that future development of the game could potentially help them. 2 participants strongly agreed and 2 participants agreed that they were proficient at playing games and that the game provided a good overview of requirements collection and analysis. All participants agreed that the realism in the game was adequate and that GBL could be used to increase skills at HE level.

4.6.3.2. Control group (role-play). The students were generally supportive of the role-play in this study. 3 of the 4 participants agreed that they would be willing to participate in the role-play over a prolonged period of time, that they would participant in the role-play as part of a computing course, that the role-play was engaging and that it sustained their engagement. 3 of the participants agreed that they would like to participate in a similar activity again. 1 strongly agreed and 2 agreed that the role-play fitted well into its intended context and 2 agreed that it could be used in its anticipated domain in the future. 3/4 participants believed that the advice in the role-play was adequate, however 1 strongly agreed and 2 agreed that the role-play required increased realism to improve it. 2 agreed that future development of the role-play could potentially help learners of their demographic and 3 agreed that future development of the role-play could potentially help them. 1 strongly agreed and 2 agreed that they were proficient at role-playing and 2 agreed that the role-play activity gave a good overview of requirements collection and analysis. 3 agreed that role-playing could help increase their skills and could be used to increase skills at HE level.

4.7. Study four in FE colleges (college A, B, C and D)

4.7.1. Learning effectiveness

A Mann–Whitney *U* test indicated that there was no significant difference in knowledge levels in the pre-test between the game and role-play groups. Participants' scores on the knowledge requirements tests before and after both interventions were compared. The mean score on the pre-test in the role-play group was 5.36 out of 13 (SD = 1.82) and the mean score on the post-test was 6.64 out of 13 (SD = 2.13). A Wilcoxon matched-pairs signed ranks test indicated that the increase in knowledge was significant ($Z = -1.836$, $p < 0.05$). The mean score in the pre-test in the game group was 4.83 (SD = 2.37) and on the post-test was 6.43 (SD = 2.04). A Wilcoxon matched-pairs signed ranks test indicated that the increase in knowledge was significant ($Z = -2.574$, $p < 0.01$). A Kruskal–Wallis test between the difference in the knowledge scores of the pre and post-test was not significant ($\chi^2 = 0.169$, $p > 0.05$). This indicates that the variation of score-difference within groups is larger than the differences between the groups. This means that while both groups increased their knowledge, no one group significantly outperformed the other.

4.7.2. Aspects ratings

4.7.2.1. Experimental group (game). The ratings of the aspects of the game by the FE students were generally positive. Table 8 shows the ranking of the aspects of the game for the FE students ranked from highest (5) to lowest (1).

4.7.2.2. Control group (role-play). The ratings of the aspects of the role-play by the FE students were generally positive Table 9 shows the ranking of the aspects of the role-play for the FE students ranked from highest (5) to lowest (1).

Table 6
Ratings of the aspects of the game.

Aspect	Rank	Mean	SD
Control mechanism and interface	1st	4.50	0.58
Navigation inside the environment	2nd	4.25	0.5
Character customisation	3rd	4.00	0.00
Graphics	4th	3.75	0.5
Help and scaffolding	4th	3.75	0.5
Realism of the environment	4th	3.75	0.5
Realism of the characters in the environment	4th	3.75	0.5
Narrative and dialog	5th	3.25	0.96
Clear goal structure	6th	3.00	1.41
Collaboration	7th	2.5	1.73
Sound	8th	0.75	0.96

Table 7
Ratings of the aspects of the role-play.

Aspect	Rank	Mean	SD
Collaboration	1st	3.75	0.5
Ability to improvise	1st	3.75	0.5
Narrative and dialog	2nd	3.5	0.58
Clear goal structure	3rd	3.25	0.96
Help and scaffolding	4th	2.5	0.58
Realism of the characters in the dialog	4th	2.5	0.58
Realism of the scenario	5th	2.25	0.5

4.7.3. Learner perceptions

4.7.3.1. Experimental group (game). The FE student perceptions of the game were not particularly positive. 8 out of 23 agreed that they would be willing to play the game as part of a computing course and 6 agreed that they would be willing to play the game over a prolonged period of time. 1 strongly agreed and 5 agreed that the game was engaging while 2 strongly agreed and 3 agreed that the game sustained their engagement. 1 participant strongly agreed and 1 agreed that they would like to play the game again. 2 participants strongly agreed and 13 agreed that the game fitted well into its intended context while 1 strongly agreed and 9 agreed that the game could be used in its anticipated domain in the future. 3 strongly agreed and 11 agreed that the advice in the game was acceptable 2 strongly agreed and 10 agreed that future development of the game could potentially help learners of their demographic and that future development of the game could potentially help them. 1 strongly agreed and 10 agreed that the game provided a good overview of requirements collection and analysis. 9 participants believed that the realism of the game was adequate, 2 strongly agreed and 5 agreed that the game could increase their skills and 1 strongly agreed and 10 agreed that GBL could be used to increase skills at FE level.

4.7.3.2. Control group (role-play). The FE student perceptions of the role-play activity were generally positive. 6 out of the 14 participants agreed that they would be willing to participant in the role-play over a prolonged period of time, 1 strongly agreed and 6 agreed that they would be willing to participate in the role-play activity as part of a computing course, however only 5 agreed that they would like to participate in a similar activity. 9 participants agreed that the role-play was engaging and 1 strongly agreed and 8 agreed that the activity sustained their engagement. 1 strongly agreed and 9 agreed that the activity fitted well into its intended context and that the activity could be used in its anticipated domain in the future. 1 strongly agreed and 7 agreed that the advice in the role-play was adequate. 1 strongly agreed and 9 agreed that future development of the activity could potentially help learners of their demographic and 8 agreed and 1 strongly agreed that future development of the role-play activity could potentially help them. 10 participants believed that the role-play exercise provided a good overview of requirements collection and analysis and that it could increase their skills. 3 strongly agreed and 7 agreed that they believed the role-play could be used to increase their skills at FE level. 10 agreed that the role-play exercise required increased realism to improve it and only 5 agreed that the realism of the role-play exercise was adequate. 1 participant strongly agreed and 2 agreed that they were proficient at role-playing activities.

4.8. Study five comparison of HE and FE students

Overall 24 students evaluated the game at HE level and 23 evaluated the game at FE level. 21 participants were in the role-play group at HE level and 14 were in the role-play group at FE level. In terms of knowledge acquisition in the two game groups, due to the addition of questions as a result of the pilot study, the two game groups can be compared in two different ways. Firstly the 24 students in HE can be compared to the 23 in FE by taking the 6 original knowledge questions into consideration and secondly the 8 students at HE can be compared to the 23 students at FE by taking all of the additional questions into account, including the original 6.

4.8.1. Game Group using original six knowledge questions

When all of the HE students are taken into account in the game group a Wilcoxon matched-pairs signed ranks test indicated that the increase in knowledge between the pre and post-tests was significant ($Z = -2.610$, $p < 0.01$) indicating that the RCAG was a suitable supplementary learning tool at HE level to teach requirements collection and analysis concepts. When all of the FE students are taken into account in the game group a Wilcoxon matched-pairs signed ranks test indicated that the increase in knowledge was significant which also shows that there is an increase in knowledge between the pre and post tests ($Z = -2.371$, $p < 0.01$) indicating that the RCAG was a suitable supplementary learning tool at FE level. A Mann–Whitney *U* test was conducted to address whether there was a significant difference in initial

Table 8
Ratings of the aspects of the game.

Aspect	Rank	Mean	SD
Navigation inside the environment	1st	3.83	0.83
Graphics	2nd	3.78	0.85
Realism of the environment	3rd	3.65	0.88
Control mechanism and interface	4th	3.48	0.85
Realism of the characters in the environment	5th	3.35	1.03
Help and scaffolding	6th	3.30	0.88
Clear goal structure	6th	3.30	1.18
Narrative and dialog	7th	3.17	1.40
Character customisation	8th	2.82	1.43
Collaboration	9th	2.50	1.71
Sounds	10th	0.87	1.42

Table 9
Ratings of the aspects of the game.

Aspect	Rank	Mean	SD
Narrative and dialog	1st	3.64	0.74
Clear goal structure	2nd	3.54	0.88
Ability to improvise	3rd	3.50	0.90
Realism of scenario	4th	3.43	1.02
Realism of characters in the dialog	5th	3.29	1.20
Help and scaffolding	6th	3.07	1.07
Collaboration	6th	3.07	0.83

knowledge in the pre-test between HE and FE students ($Z = -2.741$, $p < 0.01$). Students at HE level had a significantly higher level of knowledge scoring higher in the pre-test (4.04, SD = 1.55) than FE students (3.70, SD = 1.43). On completion of the post-test, while both groups had increased their knowledge scores, a Mann–Whitney *U* test revealed that there was still a significant difference in knowledge between FE and HE post-test scores ($Z = -3.471$, $p < 0.001$) and students at HE level still had a higher level of knowledge than FE students. This indicates that while the RCAG is providing a suitable supplementary learning experience, it may be more suited to HE level in the current version and certainly raises the issue that the different levels of knowledge in FE and HE students should be considered. It may be a worthwhile consideration to produce a simplified version of the RCAG to address lower levels of initial knowledge at FE level in future research.

4.8.2. Game group using additional knowledge questions

A comparison of the FE and HE students using the 8 students from HE and the 23 students from FE in the game group with the additional questions was performed. A Mann–Whitney *U* test revealed that there was no significant difference between the levels of knowledge of the HE and FE students in the pre-test ($Z = -0.728$, $p > 0.05$). However in the post-test it was noted that the scores of the HE students (10.50, SD = 1.60) was significantly higher than the students in FE (6.43, SD = 2.04) ($Z = -3.798$, $p < 0.001$). Again this tends to indicate that the current version of the game is more appropriately suited to students in HE. FE students could possibly benefit from a simpler version of the game in terms of difficulty of the case study. A Kruskal–Wallis test between the difference in the knowledge scores of the FE and HE groups with the additional knowledge questions was significant ($\chi^2 = 3.551$, $p < 0.05$). The mean rank for HE students was 21.19 and for FE students 14.20 meaning that in this particular case the HE students outperformed the FE students.

4.8.3. Role-play Group using original six knowledge questions

A comparison of the role-play groups at HE and FE with the 6 original questions was performed. 21 students participated at HE level and 14 participated at FE level. A Mann–Whitney *U* test revealed that there was a significant difference in the level of knowledge between the two groups in the pre-test ($Z = -1.893$, $p < 0.05$). The HE students had a significantly higher level of starting knowledge with a mean score of 3.86 out of 6 (SD = 0.91), the FE students had a mean score of 3.07 (SD = 1.38). There was also a significant difference between the level of knowledge of the two groups in the post-test ($Z = -2.195$, $p < 0.05$) with the HE group scoring significantly higher (4.80, SD = 1.08) than the FE group (3.78, SD = 1.42). A Kruskal–Wallis test between the difference in the knowledge scores of the FE and HE groups in the role-play with the 6 original knowledge questions was not significant ($\chi^2 = 0.788$, $p > 0.05$). This means that the one group did not do significantly better or worse than the other.

4.8.4. Role-play Group using additional knowledge questions

A comparison of the FE and HE students using the 8 students from HE and the 14 students from FE in the role-play group with the additional questions was performed. A Mann–Whitney *U* test revealed that there was a significant difference between the levels of knowledge of the HE and FE students in the pre-test ($Z = -1.950$, $p < 0.05$). HE students possessed a higher level of initial knowledge in the pre-test (6.50, SD = 1.20) than FE students (5.36, SD = 1.13). In the post-test it was noted that the scores of the HE students (9.125, SD = 1.13) was significantly higher than the students in FE (6.64, SD = 2.13) ($Z = -2.685$, $p < 0.01$). This indicates that the role-play exercise, like the game intervention may be more suited to HE students than FE. A Kruskal–Wallis test between the difference in the knowledge scores of the FE and HE groups in the role-play with the additional knowledge questions was not significant ($\chi^2 = 1.663$, $p > 0.05$). This means that the one group did not do significantly better or worse than the other.

4.8.5. Comparison of aspects ratings

4.8.5.1. Game aspects. Table 10 shows the comparison of the rankings of the aspects of the game of FE and HE learners.

Despite a difference in the ranking of the aspects between FE and HE learners, a Mann–Whitney *U* test revealed that this was not significant. This indicates that the aspects of the game did not strongly influence learning in either the FE or the HE group. The students at HE level seemed to appreciate the character customisation facilities to a greater extent.

In terms of aspects of the game the pre-test asked the participants to rate the importance of the aspects of games in general and the post-test asked them to rate these aspects in the RCAG. This means that whether the aspects of the game have met the expectations of the two learner groups can be assessed. With the FE students a Wilcoxon matched-pairs signed ranks test indicated that the game met the expectations with regards to graphics, help and scaffolding facilities, and realism of the environment. The game did not meet the expectations of the learners at FE level with regards to the following aspects: narrative and dialog, character customisation, navigation inside the environment, realism of the characters in the environment, control mechanism and interface and clear goal structure. The mean rating for these aspects in the post-test were significantly lower than in the pre-test indicating that these are areas that may require improvement in the game.

With the HE students a Wilcoxon paired signs ranked test indicated that the game met the majority of the expectations of the learners. There was no significant difference between pre and post-test with regards to the following 6 aspects: graphics, narrative and dialog, help and scaffolding facilities, character customisation, realism of the environment, and control mechanism and interface. The results indicate

Table 10
Comparison of the aspects of the game between HE and FE.

Aspect	FE Level			HE Level		
	Rank	Mean	SD	Rank	Mean	SD
Navigation inside the environment	1st	3.83	0.83	2nd	3.92	0.83
Graphics	2nd	3.78	0.85	3rd	3.79	0.78
Realism of environment	3rd	3.65	0.88	2nd	3.92	0.072
Control mechanism and interface	4th	3.48	0.85	1st	4.13	0.83
Realism of the characters in the environment	5th	3.35	1.03	5th	3.50	0.88
Help and scaffolding	6th	3.30	0.88	8th	3.29	0.81
Clear goal structure	6th	3.30	1.18	6th	3.38	1.10
Narrative and dialog	7th	3.17	1.40	7th	3.33	0.96
Character customisation	8th	2.83	1.43	4th	3.54	0.78
Collaboration	9th	2.50	1.71	9th	2.17	1.71

that the game did not meet the HE learners' expectations with regards to: navigation inside the environment, realism of the characters in the environment and clear goal structure.

4.8.5.2. Role-play aspects. Table 11 shows the comparison of the rankings of the aspects of the role-play activity of FE and HE learners.

Despite a difference in the ranking of the role-play aspects, a Mann Whitney *U* test indicated that there were no significant differences in any of the aspects of the role-play exercise between FE and HE. This indicates that the aspects of the role-play exercise did not strongly influence learning in both FE and HE.

The expectations of the aspects of the role-playing exercise can also be assessed to address whether the actual exercise met the expectations of the learners. With the FE students a Wilcoxon matched-pairs signed ranks test indicated that the role-play exercise met the expectations of the learners in every category. This was not the case at HE level as a Wilcoxon matched-pairs signed ranks test indicated the role-play exercise failed to meet the expectations of the learners in every category.

4.8.6. Comparison of learner perceptions

4.8.6.1. Game groups. Mann–Whitney *U* tests indicated that overall HE students were more accepting of the game. HE students were significantly more willing to play the game over a prolonged period of time ($Z = -3.145$, $p < 0.001$) (Mean = 3.33 out of 5, SD = 1.09) than FE students (Mean = 2.13, SD = 1.29). They were also more willing to play the game as part of a computing course ($Z = -3.866$, $p < 0.001$) (HE Mean = 4.00, SD = 0.59; FE Mean = 2.78, SD = 1.13) and found the game more engaging ($Z = -3.399$, $p < 0.001$). HE students also expressed a significantly greater desire to play the game again ($Z = -3.790$, $p < 0.001$) and felt that it could be used in its anticipated domain in the future ($Z = -3.425$, $p < 0.001$) and at tertiary education level ($Z = -2.661$, $p < 0.01$). HE students were also significantly less distracted while playing the game (Mean = 2.17, SD = 1.17) than FE students (Mean = 3.09, SD = 1.28) ($Z = -2.621$, $p < 0.01$) and believed that future development of the game could potentially help people of their demographic more significantly than FE students (HE Mean = 3.88, SD = 0.99; FE Mean = 3.04, SD = 1.49) ($Z = -2.216$, $p < 0.01$). HE students also more firmly believed that the game provided a good overview of requirements collection and analysis ($Z = -2.217$, $p < 0.01$).

Both groups believed that the advice quality in the game was acceptable ($Z = -0.451$, $p > 0.05$) (HE Mean = 3.54, SD = 0.93; FE Mean = 3.57, SD = 1.08), and that the game was complex ($Z = -0.044$, $p > 0.05$). Both groups believed that they were proficient at playing games (HE Mean = 4.00, SD = 1.14; FE Mean = 3.73, SD = 1.21) and believed that the realism of the game was adequate and that generally the game did not require increased or decreased realism to improve it. Both groups also believed that future development of the game could potentially help them (HE Mean = 3.87, SD = 0.99; FE Mean = 3.04, SD = 1.49); ($Z = -0.997$, $p > 0.05$).

Participants were asked the open ended question of what they would like to see improved in the game. Some of the more interesting suggestions were:

- "Graphics, sounds added", "better display of text, better instructions", "user interface/dialogue. An option to view the requirements on a larger screen. Dialogue tidied up and allowed scrolling." "More staff members, bigger office."

4.8.6.2. Role-play groups. A Mann–Whitney *U* test showed no significant differences in perceptions of the role-play activity between the HE and FE groups. It is possible that the differences in perceptions in the game groups is related to the fact that GBL is a newer, more novel approach and role-play is in some respects more familiar. Participants were also asked the open ended question of what they would like to see improved in the role-play activity. Some of the more interesting suggestions were:

Table 11
Comparison of role-play activity between HE and FE.

Aspect	FE Level			HE Level		
	Rank	Mean	SD	Rank	Mean	SD
Narrative and dialog	1st	3.64	0.74	2nd	3.38	0.80
Clear goal structure	2nd	3.54	0.88	1st	3.43	1.03
Ability to improvise	3rd	3.50	0.90	2nd	3.38	0.80
Realism of scenario	4th	3.43	1.02	3rd	3.05	0.92
Realism of the characters in the dialog	5th	3.29	1.20	4th	2.76	0.99
Help and scaffolding facilities	6th	3.07	1.07	5th	2.57	1.08
Collaboration	6th	3.07	0.82	2nd	3.38	1.12

- “A more detailed list of questions”, better “flow”, better “improvisation”, including “more background info”, “more realistic environment”.

5. Discussion

Overall the studies performed have shown that GBL can be a suitable approach to teach requirements collection and analysis at a supplementary level in tertiary education. While the control and experimental groups all experienced a significant increase in knowledge at both FE and HE level, the research performed indicated that a GBL approach to teach requirements collection and analysis may be more suitable to HE level than FE level. With regards to the original six knowledge questions, the HE experimental group learners had a higher level of knowledge in the pre-test and a higher level of knowledge on the post-test in comparison with the FE group experimental group learners. This suggests that different levels of initial knowledge should be taken into account when considering a GBL approach for different educational levels. When comparing the HE and FE students with the additional questions there was no significant difference between knowledge levels on the pre-test, however, HE students had a significantly higher level of knowledge in the post-test again indicating that the GBL approach is more effective at HE level than FE level. There was no significant difference in the ratings of the aspects of the game between the groups indicating that the aspects remained consistent and did not aversely affect the results.

The results indicate that the game did not meet the expectations of FE learners to the same degree as the HE learners. The game met FE expectations with regards to 3 aspects; however it met HE expectations with regards to 6 aspects. Interestingly the role-play activity met the expectations of FE students but did not meet any of the expectations of any of the categories in HE. This is possibly because role-play is a better known technique at FE level and GBL is a newer, more novel technique. The differences in expectations could not be explained by a difference of age or any significant difference in expectation levels in the pre-test between the two groups.

In terms of perceptions HE students were more accepting of the game. They were more willing to play the game over time, play the game as part of a computing course and found the game more engaging. HE students also expressed a greater desire to play the game again and felt that the game could be used in its anticipated domain in the future. HE students were less distracted while playing the game and believed that future development of the game could potentially help people of in HE more.

Both groups believed that the game fitted well with its intended context, that the advice in the game was acceptable and that the game was complex. They also believed that the realism of the game was adequate, that the game did not require increased or decreased realism to improve it and that future development of the game could potentially help them.

In conclusion the studies showed that a GBL approach can be used to teach software engineering concepts at a supplementary level in tertiary education. The approach seems to be more suited to HE learners in terms of knowledge acquisition, aspects of the game and perceptions as opposed to FE learners. Despite the fact that the game did not meet the expectations of FE learners and was perceived slightly more negatively, the game proved to be just as effective as role-play in providing a supplementary learning experience at FE level. One possible explanation for this difference in perceptions and expectations is differing levels of maturity between the groups of learners.

This research has shown that a game can be just as effective as role-playing and more effective than paper-based case studies at teaching requirements collection and analysis in tertiary education. Hays (2005), however, raises two issues about the generalisability of GBL: although games can be effective for learning for different tasks, this does not inform us whether to use a game to teach a specific instructional task and that we should not generalise on the effectiveness of one game for a group of learners for a particular task to all games for all learners for all tasks. Generalisability is an issue of great concern for the future of GBL as it is impossible to take one game and apply it to one area and then proceed to make wider generalisations. While this paper has provided a preliminary evaluation that has shown the game can be used to teach requirements collection and analysis, before the game can be used effectively in a course we believe more substantial qualitative study and experience is required.

5.1. Comparison with other studies

Connolly et al. (2007) performed a literature search to identify empirical evidence associated with the use of computer games for learning in computer science, information systems and software engineering. Two more extensive literature reviews were performed in 2008 and 2009: one focusing on developing an evaluation framework for GBL (Connolly et al., 2009) and another focusing on learning value and skill enhancement of gaming and methods of measuring the resultant outcomes and impacts (Boyle et al., submitted for publication). Out of the 24 studies identified, only three (13%) used a larger sample size with an appropriate method of control. Two of these studies were attempting to assess whether a GBL approach outperformed a traditional approach. No study focused on requirements collection and analysis in software engineering education. This paper has contributed the first detailed, preliminary empirical evaluation of a GBL application to teach requirements collection and analysis in software engineering education that has been compared against traditional teaching approaches.

6. Future research

In terms of future research directions, we hope to perform further quantitative evaluations at different institutions at different educational levels to produce further empirical evidence associated with the game. We also hope to further develop the game and evaluate learning in different roles such as the project manager, systems designer or team leader After the game has been sufficiently tested, we hope to incorporate it more formally into our computing curriculum.

GBL requires more extensive empirical evidence if it is to become a recognised teaching approach. To assist researchers in gathering empirical evidence by performing GBL evaluations, Hainey, Connolly and Boyle (2010) have recently produced a refined evaluation framework based on the one used in this paper. The refined evaluation framework is designed to be a starting point for researchers to focus the evaluation of a GBL application. The following guidelines may be useful in carrying out a GBL evaluation:

1. Formulate the research questions for the study of the GBL application.

2. Based on the research questions, produce a shortlist of the main things to evaluate about the GBL application, for example, learning effectiveness, how motivated the learners are to participate and what particular aspects are interesting to evaluate.
3. Go through all of the measurements associated with each framework category (i.e. learner performance, GBL environment, learner/instructor motivation, learner/instructor perception, learner instructor preferences and collaboration) and identify those measurements that are of particular interest.
4. Examine the literature associated with each relevant measurement and check how the identified empirical studies have collected evaluation data on this measurement.
5. When all of the relevant measurements have been identified, select an appropriate experimental design to properly collect these measurements; for example, pre-test/post-test or pre-test/post-test experimental control group design.
6. Run the evaluation using the chosen experimental design methodology to address the research questions.

The evaluation framework is at this stage designed to be used in an implicit way, however, it will produce a number of starting ideas for a more focused and rigorous GBL evaluation.

References

ADPRIMA. (2009). *Instructional methods information website*. Retrieved 21 May, 2009, from. http://www.adprima.com/teachmeth.htm.

Aldrich, C. (2005). *Learning by doing: A comprehensive guide to simulations computer games, and pedagogy in e-Learning and other educational experiences*. Pfeiffer.

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. *Journal of Adolescent Health, 41*, 263–270.

Bonwell, C. C. (1996). Enhancing the lecture: revitalizing a traditional format. In T. E. Sutherland, & C. C. Bonwell (Eds.), *Using active learning in college classes: A range of options for faculty, New Directions for Teaching and Learning, No. 67* (pp. 31–44). Place: Wiley Periodicals Inc.

Boyle, E., Connolly, T. M., Mac Arthur, E., Hainey, T., Hancock, F., Boyle, J. Engagement in computer games: a systematic literature review. *International Journal of Human Computer Studies*, submitted for publication.

Bransford, J., Sherwood, R., Hasselbring, T., Kinzer, C., & Williams, S. (1990). Anchored instruction: why we need it and how technology can help. In D. Nix, & R. Spiro (Eds.), *Cognition, education, & multimedia: Exploring ideas in high technology* (pp. 163–205). Hillsdale, NJ: Lawrence Erlbaum Associates.

Briggs, J. (1994). In G. King, C. Brebbia, M. Ross, & G. Staples (Eds.), *Do students want to engineer software? Software engineering in higher education*. Southampton: Computational Mechanics Publications.

Brooks, F. (1987). No silver bullet: essence and accidents of software engineering. *IEEE Computer, 20*(4), 10–19.

Cashin, W. E. (1985)*Improving lectures idea paper, No. 14*. Manhattan: Kansas State University, Center for Faculty Evaluation and Development.

Christoph, N. (2007). *The role of metacognitive skills in learning to solve problems*. PhD Thesis submitted to the University of Amsterdam. Retrieved 26th May, 2010 from http://dare.uva.nl/document/22568.

Chua, A. Y. K. (2005). The design and implementation of a simulation game for teaching knowledge management. *Journal of American Society for Information Science and Technology, 56*(11), 120–1216.

Connolly, T. M., Stansfield, M. H., & Hainey, T. (2007). An application of games based learning within software engineering. *British Journal of Educational Technology., 38*(3), 416–428.

Connolly, T. M., Stansfield, M. H., & Hainey, T. (2009). Towards the development of a games-based learning evaluation framework. In T. M. Connolly, M. H. Stansfield, & E. Boyle (Eds.), *Games-based learning advancement for multisensory human computer interfaces: Techniques and effective practices*. Idea-Group Publishing: Hershey, ISBN 978-1-60566-360-9.

Connolly, T. M., Stansfield, M. H., McLellan, E., Ramsay, J., & Sutherland, J. (November 2004). Applying computer games concepts to teaching database analysis and design. In *Proceedings of the International conference on computer games*. Reading, UK: AI, design and education.

Davis, B. G. (2001). *Tools for teaching*. San Francisco, CA: Jossey-Bass Publishers.

Dawson, R. (2000). Twenty dirty tricks to train software engineers. In *Proceedings of the 22nd international conference on software engineering* (pp. 208–209). Limerick, Ireland: ACM Press.

de Freitas, S. (2006). *Learning in immersive worlds*. Joint Information Systems Committee.

Ford, C. W., & Minsker, S. (2003). TREEZ – an educational data structures game. *Journal of Computing Sciences in Colleges, 18*(6), 180–185.

Habgood, M. P. J. (2007). The effective integration of digital games and learning content. Thesis submitted to the University of Nottingham. Retrieved 27th October, 2008 from http://zombiedivision.co.uk/.

Hays, R. T. (2005). The effectiveness of instructional games: a literature review and discussion. Retrieved September 6, 2009, from. http://adlcommunity.net/file.php/36/GrooveFiles/Instr_Game_Review_Tr_2005.pdf.

Hainey, T., Connolly, T. M., & Boyle, L. (20–22 October, 2010). A refined evaluation framework for game-based learning. In: *4th European Conference on Games-based Learning (ECGBL)*. Copenhagen, Denmark.

Healy, A. & Connolly, T. M. (25–26 October, 2007). Does games-based learning, based on a constructivist pedagogy, enhance the learning experience and outcomes for the student compared to a traditional didactic pedagogy? In: *Proceedings of 1st European Conference on Games-Based Learning. (ECGBL)*. Paisley, Scotland.

Jain, A., & Boehm, B. (2006). SimVBSE: Developing a game for value-based software engineering. In: *Proceedings of 19th Conference on Software Engineering Education and Training (CSEET)*. (pp. 103–114). Turtle Bay Resort, Oahu, Hawaii.

Johnson, W. L. & Wu, S. (April, 2008) Assessing aptitude for learning with a serious game for foreign language and culture. In: Proceedings of 9th international conference on intelligent tutoring systems. Montreal.

Kolb, D. (1984). *Experiential learning*. New Jersey: Prentice-Hall Inc.

Laurillard, D. (June 1996). Keynote speech at the ACM SIGCSE/SIGCUE conference. Barcelona.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, England: Cambridge University Press.

Lennon, J. L. (2006). Debriefings of web-based malaria games. *Simulation & Gaming., 37*(3), 350–356.

Ludi, S. & Collofello, J. (2001). An analysis of the gap between the knowledge and skills learned in academic software engineering course projects and those required in real: projects. In: *Frontiers in education conference*, *31st Annual* 1(1).

Nelson, B. C. (2007). Exploring the use of individualized reflective guidance in and educational multiuser virtual environment. *Journal of Science Education and Technology., 16*, 83–97.

Oh Navarro, E., Baker, A. & Van der Hoek, A. (2004). Teaching Software Engineering Using Simulation Games. In: *Proceedings of the 2004 international conference on simulation in education*. San Diego, California, January 2003 (to appear).

Oh Navarro, E. (2006). *SimSE: A software engineering simulation environment for software process education*. Unpublished doctoral thesis. California: University of Irvine.

Oh Navarro, E. & Van der Hoek, A. (2005). Design and evaluation of an educational software process simulation environment and associated model. In: *Proceedings of the eighteenth conference on software engineering education and training*. Ottawa, Canada.

Rankin, Y., Gold, R. and Gooch B. (2006). Gaming as a language learning tool. In: *Proceedings of the ACM SIGGRAPH educators program*.

Roubidoux, M. A. (2005). Breast cancer detective: a computer game to teach breast cancer screening to native American patients. *Journal of Cancer Education, 20*(1), 87–91.

Schneider, M., Carley, K., & Moon, I. (2005). *Detailed comparison of America's army game and unit of action experiments*. Carnegie Mellon University, School of Computer Science, Institute for Software Research International. (Technical Report CMU-ISRI-05-139).

Shaffer, D. W., Squire, K. T., Halverson, R., & Gee, J. P. (2004). *Video games and the future of learning*. Phi Delta Kappan. http://www.academiccolab.org/resources/gappspaper1.pdf Retrieved 6th December, 2008 from.

Shaw, K. & Dermoudy, J. (2005). Engendering an empathy for software engineering. In: *Proceedings of the 7th Australasian computing education conference (ACE2005), Vol. 42*. (pp. 135–144). Newcastle, Australia.

Sommerville, I. (2006). *Software engineering* (8th ed.). International Computer Science.

Squire, K., Barnett, B., Grant, J. M., & Higginbotham, T. (2004). Electromagnetism Supercharged! Learning physics with digital simulation games. In: *Proceedings of the international conference on learning sciences, Vol. 6*, (pp. 513–520).

Waraich, A. (2004). Using narrative as a motivating device to teach binary arithmetic and logic gates. In: *Proceedings of the 9th annual SIGCSE conference on innovation and technology in computer science education*. (pp. 97–101). Leeds, United Kingdom.

Wehrli, G. & Nyquist, J. G. (2003). Creating an educational curriculum for learners at any level. In: *Proceedings of the AABB conference*. Retrieved May 27, 2009, from http://www.nhchc.org/UNMSOM/ResourcesforDevCurr_Teaching_LearningActivit ies.pdf.

Young, J., & Upitis, R. (1999). The microworld of Phoenix Quest: social and cognitive considerations. *Education and Information Technologies, 4*(4), 391–408.

Zhu, Q., Wang, T., & Tan, S. (2007). Adapting game technology to support software engineering process teaching: from SimSE to Mo-SEProcess. In: *Proceedings of third International Conference on Natural Computation (ICNC)*. (pp. 777–780).