

# Accepted Manuscript

A systematic literature review of games-based learning empirical evidence in primary education

Thomas Hainey, Thomas M. Connolly, Elizabeth A. Boyle, Amanda Wilson, Aisyah Razak



PII: S0360-1315(16)30156-7

DOI: [10.1016/j.compedu.2016.09.001](https://doi.org/10.1016/j.compedu.2016.09.001)

Reference: CAE 3061

To appear in: *Computers & Education*

Received Date: 18 January 2016

Revised Date: 2 May 2016

Accepted Date: 8 September 2016

Please cite this article as: Hainey T., Connolly T.M., Boyle E.A., Wilson A. & Razak A., A systematic literature review of games-based learning empirical evidence in primary education, *Computers & Education* (2016), doi: [10.1016/j.compedu.2016.09.001](https://doi.org/10.1016/j.compedu.2016.09.001).

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**A Systematic Literature Review of Games-based Learning Empirical Evidence in Primary Education**

### **Abstract**

*Games-based Learning (GBL) has developed a reputation with educationalists it is perceived as a potentially engaging form of supplementary learning that could enhance the educational process and has been used at all levels of education including primary, secondary and tertiary education. Despite this recognition and utilisation there is still a lack of empirical evidence supporting GBL as an approach. This paper presents the findings of a systematic literature review performed from 2000 to 2013 specifically looking at quality empirical studies associated with the application of GBL in Primary Education (PE) categorising studies into: behavioural change, affective and motivational outcomes, perceptual and cognitive skills and knowledge acquisition and content understanding. This paper presents a synthesis of these high quality studies associated with GBL. The studies showed that GBL have been used to teach a variety of subjects to children and young people in PE with mathematics, science, language and social studies being the most popular. However, the analysis shows that more Randomised Controlled Trial (RCT) studies should be performed comparing GBL to traditional teaching approaches to ascertain if GBL is a useful, viable teaching approach at PE level; there is a distinct lack of longitudinal studies and further longitudinal studies are required; further studies are required looking at whether there are pedagogical benefits of using 2D or 3D games at PE level to assess if 3D immersive games are indeed necessary; further studies are also required to perform comparisons between single and collaborative play and to identify the pedagogical benefits.*

**Keywords:** serious games; games-based learning; systematic literature review; empirical evidence

### **1. Introduction**

Despite the increased popularity of GBL, there is a lack of empirical evidence to support its validity. Addressing concerns Connolly *et al.* (2012) undertook an extensive systematic literature review of empirical evidence about the positive impacts and outcomes of computer games and GBL with respect to learning and engagement. It examined papers published between January 2004 and February 2009 which reflected the growing interest during that time in the use of digital games for learning and behaviour change. The papers identified were very diverse with respect to the phenomena researched, the purpose of the studies, the underlying theoretical models and methodological approaches adopted. A key aim of Connolly's review was to develop a workable framework for categorising digital games and their outcomes. A multi-component analysis of games and their outcomes was advocated distinguishing games in terms of game genre, i.e. the kind of games

and associated game mechanics such as whether the games were role playing, strategy games, adventure games, simulations; the platform used or method of delivery such as computer game, video console, mobile, online; and the subject discipline or curricular areas that the game addressed. The outcomes and impacts of playing games also differed primarily with respect to the learning and behavioural outcomes of the games. Based on previous research the following categories for the classification of the learning and behavioural outcomes and impacts that games addressed were identified: knowledge acquisition /content understanding; affective and motivational; perceptual and cognitive skills; behaviour change; physiological and social/soft skills outcomes.

## 1.2 GBL in PE

Primary Education (PE) is typically the first stage of compulsory education. and according to Learning.Org, is rudimentary as it provides students with the “*fundamental skills that will be the foundation for the rest of their academic careers in the following subjects: Maths, Science, Language Art, Music, History, Geography and Music.*” A number of issues can affect the success of students at PE level including well-being, teacher quality, levels of poverty and parental support. If a student fails to grasp the basics at PE level then secondary level becomes all the more difficult as students may find it difficult to fill in the gaps left in their knowledge. Wilson, Hainey and Connolly (2013) highlight the need for novel teaching approaches including the use of computing technology and computer games to promote engagement in PE. Griffiths (2002) identifies some of the benefits of computer games education. Computer games can be used for measurement of a very wide variety of tasks, as they are adaptable and easily standardised. The participants are allowed to experience curiosity, challenge, novelty and fun, possibly stimulating learning. Children can have access to technology that may help overcome technophobia and may assist in the development of transferable IT skills.

In the attempts to address the challenge of making games for education enjoyable, yet effective, researchers and educational practitioners are increasingly turning their attention towards so-called serious games for education or games-based learning. These games differ from (1) COTS games by having education as the primary goal rather than entertainment (De Freitas, 2006) and (2) the drill-and-practice games in that their complexity approaches that of their commercial counterparts. Several serious games have been developed recently for primary and secondary education. For example, Global Conflicts: Palestine is a 3D role-playing game (RPG) that deals with the Israeli-Palestinian conflict and is based on real personal stories (Egenfeldt-Nielsen and Buch, 2006) and Europe 2045, which is designed to be a supportive educational tool for social science courses in secondary schools, attempting to familiarise players with political, economic and social issues in a united Europe and the present-day world (Brom, Sisler and Slavík, 2010).

### **1.3 Games-based Learning (GBL) and Serious Games**

There is no general consensus on the definition of ‘serious games’ or what particular games can be categorised as serious. Egenfeldt-Nielsen, Smith and Tosca (2008) differentiate between three categories of educational computer games which are: research-based games, commercial entertainment games and edutainment. Commercial and research-based games encourage exploration and curiosity and go beyond edutainment games in terms of being less explicit with regards to providing the player information. Zyda (2005) defines a ‘serious game’ as: “*a mental contest, played with a computer in accordance with specific rules, that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives.*” GBL is considered to be a sub-category of serious games or can sometimes be used synonymously. For the purposes of this paper we have adopted the definition of GBL from Tang, Hanneghan and El-Rhalibi (2009) as “*the innovative learning approach derived from the use of computer games that possess educational value or different kinds of software applications that use games for learning and education purposes such as learning support, teaching enhancement, assessment and evaluation of learners.*” The majority of computer games are for the purposes of player entertainment (Dempsey *et al.*, 1996). Chen and Michael (2005) define serious games as: “*games that do not have entertainment, enjoyment or fun as their primary purpose.*” However Abt (1987, p. 9) notes that “*this does not mean that serious games are not, or should not, be entertaining.*”

To clarify the distinction between serious games, games for learning and entertainment games in this particular study: entertainment games are pre-made, COTS games that are used in PE for the purposes of learning, teaching a particular subject or promoting engagement. Games for learning or GBL in is the production of a specially implemented application for the purposes of learning, teaching a particular subject or promoting engagement. As there were only three relevant studies where one of these studies was high quality where the study referred to the application as a serious game (Wrzesien and Alcaniz Raya, 2010) this was included in the games for learning category for the purpose of analysis.

### **1.4 Research Question**

GBL has been used in a number of areas at various different levels of education and computer games are perceived as a potentially motivational approach to supplementary learning. Due to a dearth of empirical evidence supporting the validity of the approach, Connolly *et al.*, (2012) asked the research question of: “What empirical evidence is there concerning the positive impacts and outcomes of computer games?” and covered studies from 2004 - 2009. In an update to the 2012 study looking at high quality papers from 2009 to 2014, Boyle *et al.*, (2015) attempted to update Connolly *et al.*’s (2012) study in terms of bringing it up to the present. Both of these reviews looked at participants

over the age of 14 and ranged from secondary, tertiary and informal education. In an attempt to fill a gap left by these two reviews and the importance of increasing motivation and novel teaching approaches in PE, this review has focused on PE level with the following research question: “What empirical evidence is there concerning the positive impacts and outcomes of computer games at PE level?” This review will look at studies from 2000 to 2013 and will focus on the evaluation of digital computer games and will not include board games. The underlying theoretical framework has been discussed for these reviews in Connolly *et al.* (2012) and this paper has been used as a methodological reference.

## **2. Method**

### *2.1 Databases searched*

The electronic databases searched in this review included: ACM (Association for Computing Machinery), ASSIA (Applied Social Sciences Index and Abstracts), BioMed Central, Cambridge Journals Online, ChildData, Index to Theses, Oxford University Press (journals), Science Direct, EBSCO (consisting of Psychology and Behavioural Science, PsycINFO, SocINDEX, Library, Information Science and Technology Abstracts, CINAHL), ERIC (Education Resources Information Center), IngentaConnect, Infotrac (Expanded Academic ASAP), Emerald and IEEE (Institute of Electrical and Electronics Engineers) Computer Society Digital Library (CSDL).

#### *2.1.1 Search terms*

The games terms were derived from previous searches carried out on the evaluation of computer games (Connolly *et al.*, 2012) and were the following:

(“computer games” OR “video games” OR “serious games” OR “simulation games” OR  
 “games-based learning” OR MMOG OR MMORPG  
 OR MUD OR “online games”) AND (evaluation OR impacts OR outcomes OR effects OR  
 learning OR education OR skills OR behaviour OR attitude OR engagement OR motivation  
 OR affect)

#### **2.1.2 Inclusion criteria**

A number of further criteria were specified to select appropriate studies for inclusion in the review. To be included in the current review, papers had to (a) include empirical evidence relating to the impacts and outcomes of playing games. This was to address the main aim of the study in addressing the perceived lack of empirical evidence concerning digital games; (b) date from January 2000 to February 2013; (c) include an abstract and (d) include participants in PE.; e) include a focus on curricular based subjects.

## 2.2 *Quality of studies*

To assess the quality of the empirical evidence, each paper was read and given a score of 1–3 across the five dimensions used by Connolly *et al.* (2012), where 3 denotes high, 2 denotes medium and 1 denotes low on that criterion. Possible scores for each paper ranged from 5 to 15. The five dimensions were:

1. How appropriate is the research design for addressing the question, or sub-questions of this review (higher weighting for inclusion of a control group)? Papers were coded as:
  - High = 3, e.g. RCT
  - Medium = 2, e.g. quasi-experimental controlled study
  - Low = 1, e.g. case study, single subject-experimental design, pre-test/post-test design
2. How appropriate are the methods and analysis?
3. How generalisable are the findings of this study to the target population with respect to the size and representativeness of sample. To what extent would the findings be relevant across age groups, gender, ethnicity, etc.?
4. How relevant is the particular focus of the study (including conceptual focus, context, sample and measures) for addressing the question or sub-questions of this review?
5. To what extent can the study findings be trusted in answering the study question(s)?

This included scoring and specification of some of the following:

- Data Collection (information about reliability of tools/instruments, validity of data collection tools/instruments, were post-test assessments blinded)
- Data Analysis (statistical methods used, rationale, confounding factors and how they were controlled, qualitative measures)
- Results and Conclusions (summary of overall results at post and follow up study, were data from all outcomes reported, shortcomings noted by authors, implications of study)
- Quality, replicability and reporting (are the aims of the study clear, are the details of the sample used adequately described, is the data analysis adequately described, are the method adequately described).

Coding of the papers was carried out by five experienced coders who conferred about problematic coding and resolved any discrepancies. The coders included the following participants with the following qualifications:

- A Professor of Computer Games Technology with a number of publications on GBL,
- A Lecturer in Computer Games Technology with a PhD in GBL,

- A Senior Lecturer in Psychology with a number of publications in GBL,
- A Research Assistant with a PhD in GBL in PE,
- A PhD candidate studying games-based construction in PE.

All coders conferred on the classification of the papers into their various categories including game genre type. Generally when the study itself referred to a particular genre then this was the one it was placed in but multiple genre games were discussed and categorised by the 5 coders.

### *2.3 Inter-rater reliability*

To assess inter-rater reliability with respect to the quality coding of the papers, all papers were coded independently by two separate coders. The inter-rater reliability ( $r$ ) for the total scores was attained using a Cronbach Alpha test and was 0.97, showing good agreement..

## **3. Results**

### *3.1 Papers identified by search terms*

Table 1 shows the number of papers in each database identified using our search terms. The search terms identified a large number of papers (18,298) indicating the interest in the GBL area.

### *3.2 Papers selected using our inclusion criteria*

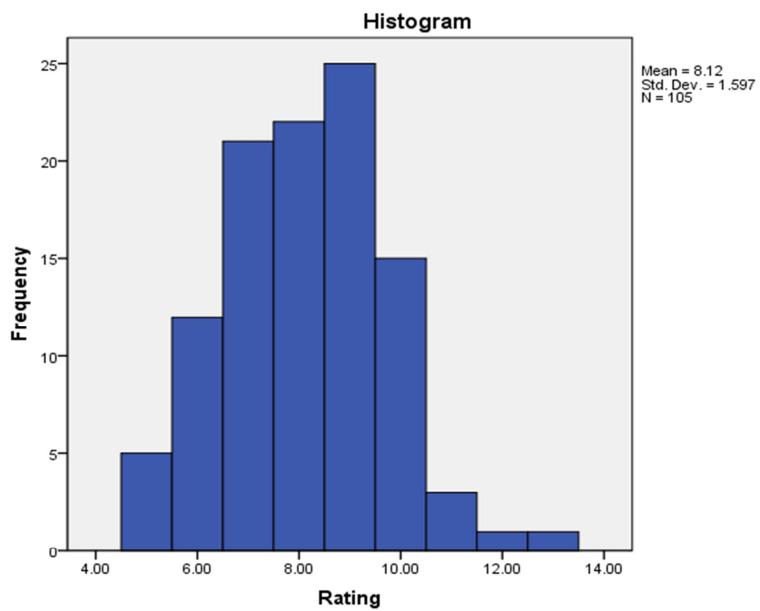
Table 1 also shows the number of papers in each database that met the inclusion criteria.

**Table 1:** Total number of papers identified from each database and number shortlisted as relevant

<i>Database</i>	Number of papers identified in the search	Number of papers meeting inclusion criteria after coding
ACM	2129	7
Biomed	59	
Cambridge	561	
Ebsco	799	2
ECGBL 2007 – 2012	303	15
ERIC	1235	9
GBL Biography	230	
IEEE (CSDL)	88	4
IEEE (Explore)	3	3
Infotrac	571	
IngentaConnect	219	
Proquest	821	2
ScienceDirect	7354	35
Web of Knowledge	1614	7
Taylor and Francis	1485	3
Wiley	657	11
Ed/IT lib	170	7
<b>Total</b>	<b>18,298</b>	<b>105</b>

### 3.3 Quality ratings

As described in Section 2.2, each of the 105 papers was given a quality score and Figure 1 shows a histogram of the number of papers rated at each score. The mean rating for the 105 papers was 8.12 ( $SD = 1.597$ ) and the modal rating for the papers was 9. Consequently papers rated 9 or over were considered as providing methodologically stronger evidence of the impact of games in that area. 45 papers were rated over 9 and these papers are summarised in Appendices A–D, showing the names of the authors, date that the study was reported, the aims and objectives of the study, methods used and the main findings and conclusions of the study. The papers were organised into the following categories: knowledge acquisition and content understanding, affective and motivational, perceptual and cognitive skill and behavioural change outcomes.



**Figure 1:** Histogram of weight of evidence (WOE) quality scores for included papers

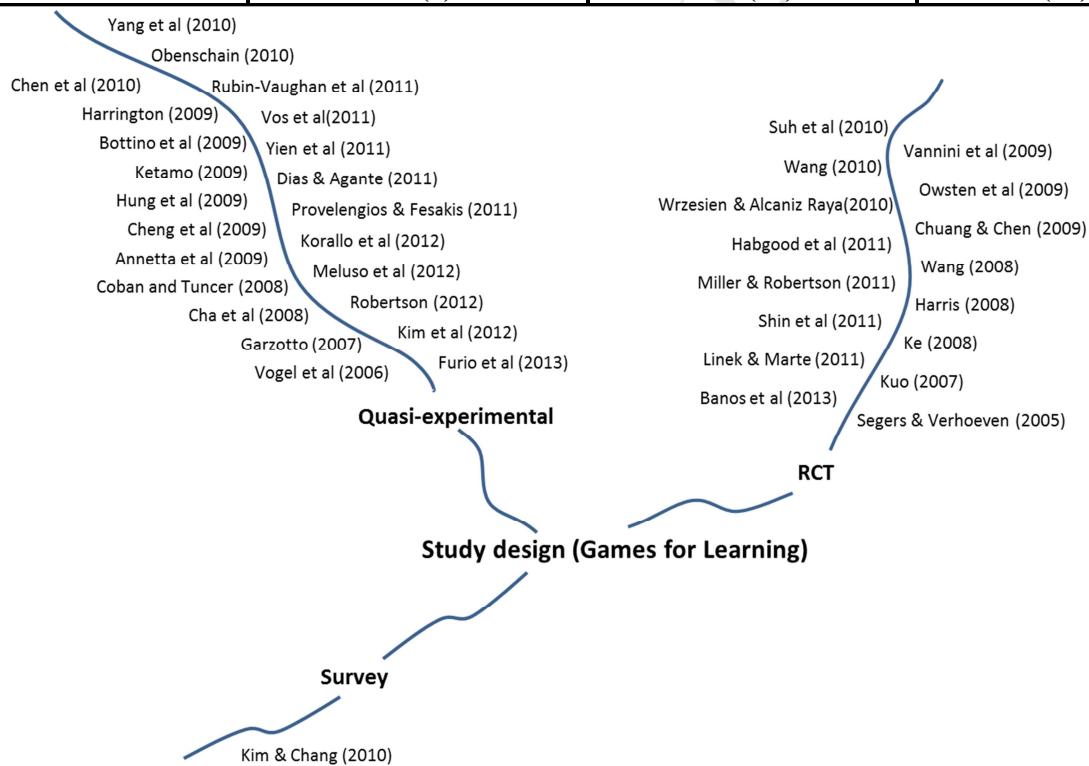
### 3.4 Study design

Table 2 shows the numbers of papers that used each study design. They are split by the primary purpose of the game which was either for learning purposes or for entertainment purposes. Results are shown both for all 105 papers with the 45 higher quality papers in brackets and this way of reporting the data is adopted in all tables. The vast majority 78 (74.3%) of the empirical papers reported quantitative data with only 27 (25.7%) reporting qualitative data. Of the 78 studies that reported quantitative data, 58 (55.2%) utilised quasi-experimental designs, 18 (17.1%) RCT designs and 2 (1.9%) used survey designs, and no studies utilised a correlational design. Figure 2 shows the 45 relevant papers in the games for learning category by study design. The results indicate a strong

preponderance towards quasi-experimental designs but interestingly unlike the previous literature review in 2012 (Connolly *et al.*, 2012) covering 14 years of age and above there was no strong use of the survey methodology this time as it is possible that this is considered and unsuitable methodology for PE. The 27 studies reporting qualitative data primarily performed case studies, observational studies during the game play or mixed methods where the number of participants was too small to yield any form of statistically significant results.

**Table 2:** Primary purpose of game (entertainment or learning) by study design (higher quality papers in brackets)

Study design	Primary purpose of Game		Total
	Entertainment	Game for Learning	
<b>Qualitative</b>	5 (0)	22 (0)	27 (0)
<b>Correlational</b>			
<b>Quasi-experimental</b>	8 (5)	50 (23)	58 (28)
<b>RCT</b>		18 (16)	18 (16)
<b>Survey</b>	1 (0)	1 (1)	2 (1)
<b>Total</b>	<b>14 (5)</b>	<b>91 (40)</b>	<b>105 (45)</b>



**Figure 2 - Study Design for Games for Learning**

### 3.5 Game variables

#### 3.5.1 The purpose of the game

Table 2 also shows the number of papers that focused on entertainment games and games for learning for all papers and for high quality papers. 14 papers looked at using entertainment

games for learning and 91 at specially implemented games for learning. The majority of studies in PE identified used a specially implemented game for learning and not a commercial off the shelf (COTS).

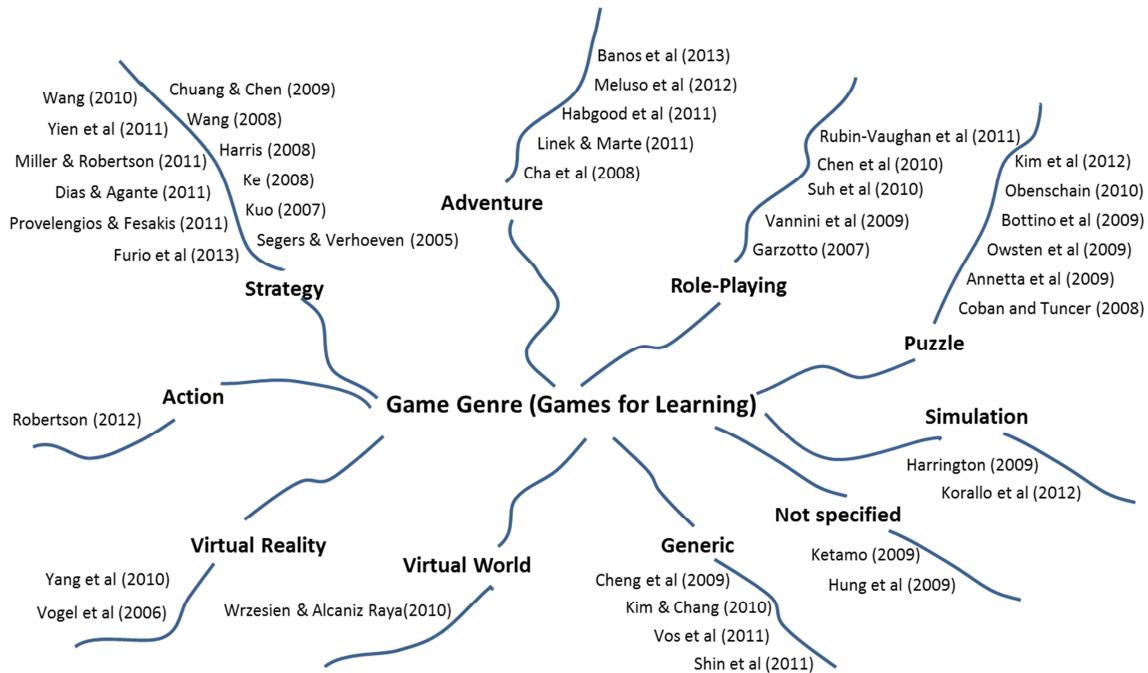
### 3.5.2 Game genre

Table 3 shows that the game genre reported in the papers in the current review were also very varied. Figure 3 shows the genres used in the games for learning category. The following genres were not used at all with regards to games for learning: fighting, platform and racing game.

Table 4 shows which designs were used in studies using different game genres. Looking at the relationship between study design and game genre reveals considerable consistency in regard to the methodology used for evaluation. For example, the outcomes of all genres of game (55%) were evaluated by quasi-experimental designs.

**Table 3:** Game genre split by the purpose of the game (higher quality papers in brackets)

Genre	Purpose of game		Total
	Entertainment	Game for Learning	
Action		2 (1)	2 (1)
Adventure	1 (0)	9 (5)	10 (5)
Animated tutorial		2 (0)	2 (0)
Generic	3 (1)	6 (4)	9 (5)
Not specified	1 (0)	7 (2)	8 (2)
Puzzle	2 (1)	16 (6)	18 (7)
Role-playing		11 (5)	11 (5)
Simulation	4 (1)	8 (2)	12 (3)
Sports		1 (0)	1 (0)
Strategy	3 (2)	24 (12)	27 (14)
Virtual reality		2 (2)	2 (2)
Virtual world		2 (1)	2 (1)
Games-based Construction Learning (GBCL)		1 (0)	1 (0)
<b>Total</b>	<b>14 (5)</b>	<b>91 (40)</b>	<b>105 (45)</b>

**Figure 3:** Genres used in the games for learning category**Table 4:** Study design by game genre (higher quality papers in brackets)

Game Genre	Study Design					Total
	Qualitative	Correlational	Quasi-experimental	RCT	Survey	
Action	1 (0)		1 (1)			2 (1)
Adventure	2 (0)		5 (2)	3 (3)		10 (5)
Animated tutorial	1 (0)		1 (0)			2 (0)
Fighting						
Generic	3 (0)		3 (3)	1 (1)	2 (1)	9 (5)
Not specified	2 (0)		5 (2)	1 (0)		8 (2)
Platform	0 (0)					
Puzzle	2 (0)		15 (6)	1 (1)		18 (7)
Racing game						
Role-playing	3 (0)		5 (3)	3 (2)		11 (5)
Simulation	6 (0)		6 (3)			12 (3)
Sports			1 (0)			1 (0)
Strategy	5 (0)		14 (6)	8 (8)		27 (14)
Virtual reality			2 (2)			2 (2)
Virtual world	1 (0)			1 (1)		2 (1)
GBCL	1 (0)					1 (0)
<b>Total</b>	<b>27 (0)</b>	<b>0 (0)</b>	<b>58 (28)</b>	<b>18 (16)</b>	<b>2 (1)</b>	<b>105 (45)</b>

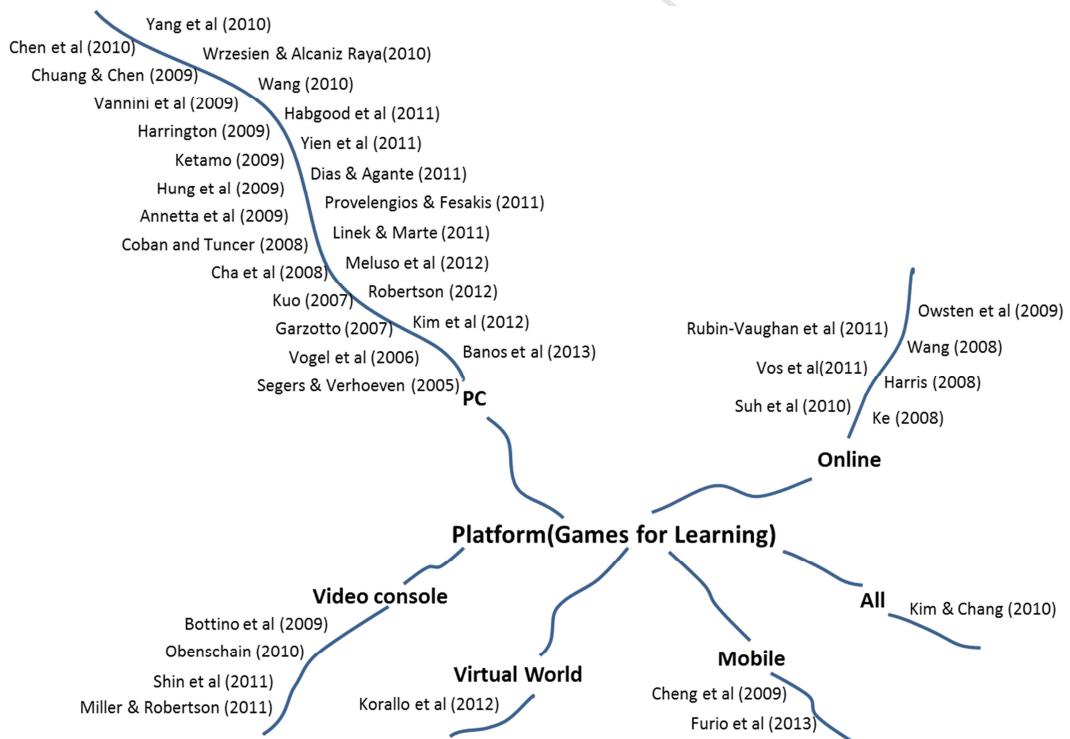
### 3.5.3 Platform delivery

Previous research noted that games for learning are delivered through a variety of platforms (Connolly *et al.*, 2012). Similar findings were evident in this study with games supporting learning

delivered via varied platforms. Table 5 shows that the most popular platforms for delivery of games was the Personal Computer (PC) (64), followed by video console (16) and online games (16), with 1 paper looking at a mobile game. The most popular platform for a customised game for learning was the PC (60) followed by online (15), video console (9) and mobile (4). The least popular platforms were virtual worlds and touch table technologies (1). Figure 4 shows the platforms used by games for learning.

**Table 5:** Purpose of game by platform (higher quality papers in brackets)

Platform	Game type		Total
	Entertainment	Game for Learning	
All	1 (0)	1 (1)	2 (1)
Mobile		4 (2)	4 (2)
Online	1 (0)	15 (7)	16 (7)
PC	4 (3)	60 (25)	64 (28)
Video console	8 (2)	9 (4)	17 (6)
Virtual world		1 (1)	1 (1)
Touch table technology		1 (0)	1 (0)
<b>Total</b>	<b>14 (5)</b>	<b>91 (40)</b>	<b>105 (45)</b>



**Figure 4:** Games for learning by platform

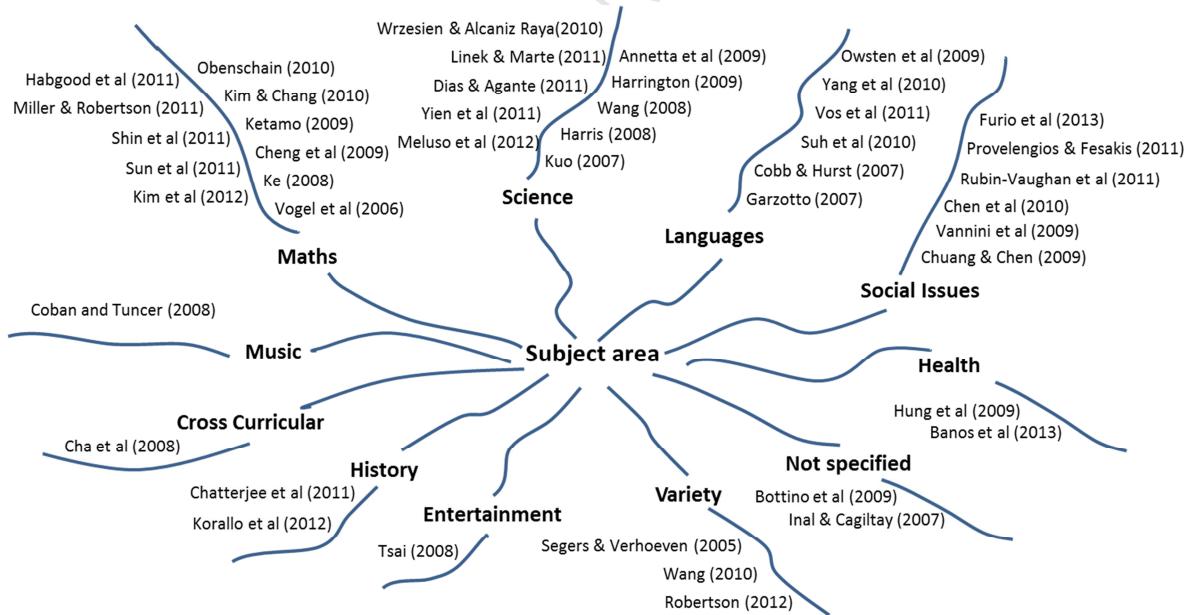
### 3.5.4 Subject discipline/curricular areas

Table 6 shows the different subject disciplines that the games described in the papers address. The games were in the areas of mathematics (25), science (21), languages (15) and social areas (13). The least popular subjects in terms of GBL being applied were personal development, music, general

knowledge and cross curricular games. It should be noted that simply because GBL has not been applied particularly frequently in a subject area that the subject is unpopular and may provide various different areas of future research application for GBL. Figure 5 shows the subject disciplines and curricular areas of the 45 relevant studies.

**Table 6:** Summary of subject discipline/curricular areas addressed by the games (higher quality papers in brackets)

Subject discipline	Total
Computing	2 (0)
Entertainment	1 (1)
General Knowledge	1 (0)
Cross curricular	1 (1)
Geography	3 (0)
Health	5 (2)
History	3 (2)
Language	15 (6)
Mathematics	25 (11)
Music	1 (1)
Not Specified	5 (2)
Personal Development	1 (0)
Science	21 (10)
Social Issues	13 (6)
Variety	8 (3)
<b>Total</b>	<b>105 (45)</b>



**Figure 5:** Subject areas/curricular disciplines of relevant studies

### 3.6 Outcomes and impacts of playing games

#### 3.6.1 Learning and behavioural outcomes of games

Table 7 shows the numbers of papers that addressed the different learning and behavioural outcomes and impacts. The most studied effects reported were knowledge acquisition and content understanding (64), affective and motivational outcomes (18), perceptual and cognitive outcomes (16), behaviour change (5) and social/soft skill outcomes (2). As expected the highest quantity of studies associated with PE were in the knowledge acquisition and content understanding category.

### *3.6.1.1 Learning and behavioural outcomes of games by study design.*

Table 7 shows that the quasi-experiment was the design of choice in although a few studies on knowledge acquisition also used RCTs (11), qualitative designs (17) and surveys (2). Quasi-experimental designs (9) were also the most frequently used when studying affective and motivational outcomes. There were also 4 RCTs (4 of high quality) and 5 qualitative studies. Quasi-experimental methodologies were also the most frequently used when studying perceptual and cognitive outcomes, behavioural change and social/soft skill outcomes.

**Table 7:** Study design by learning and behavioural outcomes of games (higher quality papers in brackets).

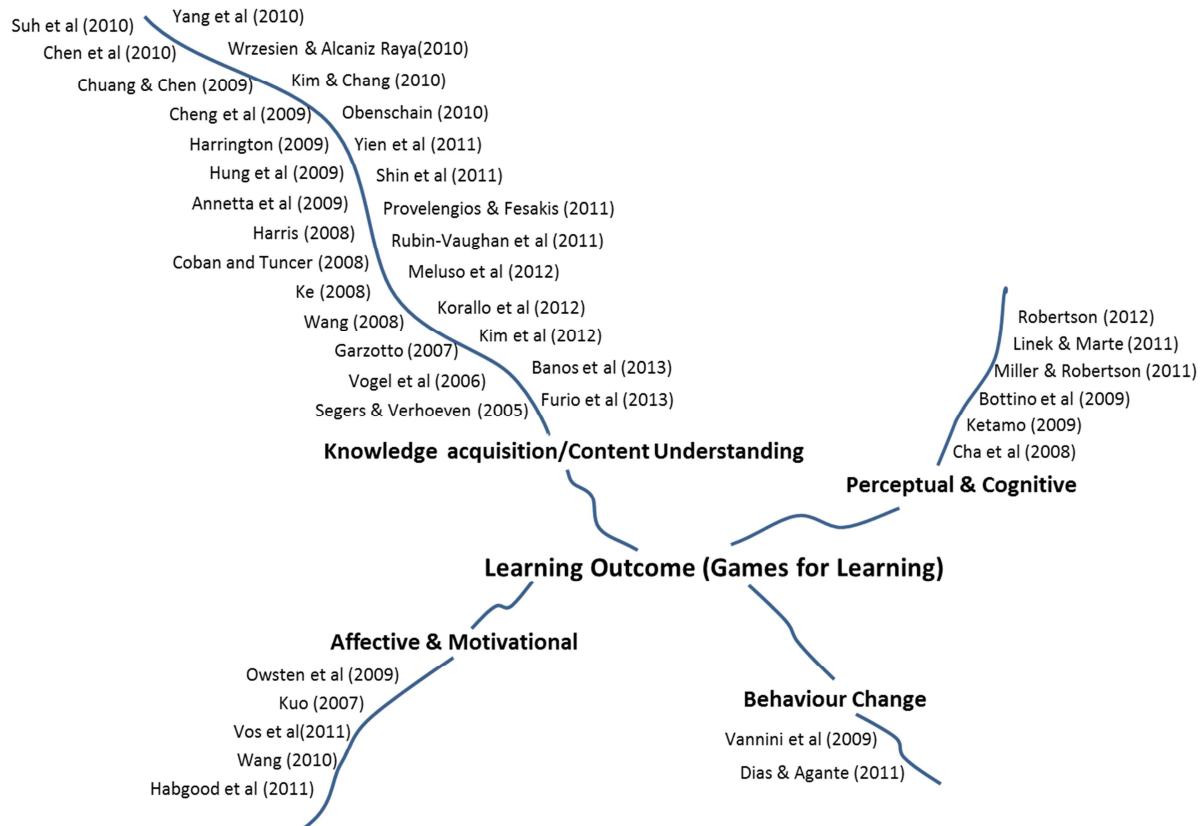
Learning outcome	Study Design					Total
	Qualitative	Correlational	Quasi-experimental	RCT	Survey	
Affective and motivational	5 (0)		9 (2)	4 (4)		18 (6)
Behaviour Change			4 (3)	1 (1)		5 (4)
Knowledge acquisition/Content understanding	17 (0)		34 (19)	11 (9)	2 (1)	64 (29)
Perceptual and cognitive	5 (0)		9 (4)	2 (2)		16 (6)
Social/soft skill outcomes			2 (0)			2 (0)
<b>Total</b>	<b>27 (0)</b>	<b>0 (0)</b>	<b>58 (28)</b>	<b>18 (16)</b>	<b>2 (1)</b>	<b>105 (45)</b>

### *3.6.1.2 Learning and behavioural outcomes by purpose of game.*

Table 8 shows the number of papers that addressed the different outcomes in terms of the purpose of the game. Once again it is evident that there are far less entertainment games used for the purposes of learning, which would suggest that games for learning in PE generally have to be specially implemented in terms of learning outcomes. The most frequently studied effects with games for learning was knowledge acquisition/content understanding (26), followed by perceptual and affective and motivational skills (14) and perceptual and cognitive skills (14). Figure 6 shows the learning outcome categorisations for games for learning. Entertainment games that were used for the purposes of learning generally followed the same trend however there were slightly more entertainment games used for the purposes of learning investigating behavioural change than games for learning.

**Table 8:** Purpose of game by learning and behavioural outcomes (higher quality papers in brackets)

Learning outcome	Game type		Total
	Entertainment	Game for Learning	
Affective and motivational	4 (1)	14 (5)	18 (6)
Behaviour Change	3 (2)	2 (2)	5 (4)
Knowledge acquisition/Content understanding	5 (2)	59 (27)	64 (29)
Perceptual and cognitive	2 (0)	14 (6)	16 (6)
Social/soft skill outcomes		2 (0)	2 (0)
<b>Total</b>	<b>14 (5)</b>	<b>91 (40)</b>	<b>(45)</b>

**Figure 6:** Learning outcome categorisations for games for learning

### *3.6.2 Discussion of the higher quality papers*

In this section the 45 higher quality papers described in Appendices A–D are discussed in more detail. Each of the categories will now be reviewed with regards to experimental design and will specifically discuss positive, negative or neutral effects, subject area, whether a 2D or 3D game was used and whether collaboration took place to identify aspects of commonality across the studies and to identify potential areas of future study. This will assist us to assess whether there is sufficient empirical evidence for the use of GBL at PE level.

#### *3.6.2.1. Knowledge acquisition/content understanding*

29 studies were found fitting into this category and the majority of the papers were looking at the acquisition of factual knowledge. These findings specifically concentrating on knowledge acquisition and content understanding have been discussed in Hainey *et al.*, (2014). In this section of the paper we are going to be looking at positivity and reliability of the overall combined results for this category to answer the question of whether there is sufficient empirical evidence to support the use of GBL in PE. RCTs are the most rigorous scientific method of hypothesis testing and way of determining cause and effect. They are the best way of demonstrating the effectiveness of novel educational approaches (Woolfson, 2011). When considering positive effects on learning and reliability, 7 RCTs reported positive learning results (Baños *et al.*, 2013, Chaung and Chen, 2009; Ke, 2008, Segers and Verhoeven, 2005; Shin *et al.*, 2011; Wang, 2008 and Suh *et al.*, 2010) where all compared a GBL approach to a traditional approach. Two of the discovered RCTs reported negative learning effects (Harris, 2008; Wrzesien and Alcaniz Raya, 2010), however Wrzesien and Alcaniz Raya (2010) reported positive motivation and attitude effects. The subject areas covered by the RCTs were three positive effect studies looking at Mathematics (Ke, 2008; Shin *et al.*, 2011; Suh *et al.*, 2010), two negative studies looking at Science (Harris, 2008; Wrzesien and Alcaniz Raya, 2010), a positive study looking at the application of GBL in Social Issues (Chaung and Chen, 2009), one in nutritional knowledge (Baños *et al.*, 2013) and two positive studies looking at a variety of subjects (Segers and Verhoeven, 2005; Wang, 2008). The results show that GBL has been more frequently applied to teaching Mathematics and not so much to teaching Science related subjects. The majority (4 studies) of GBL applications or games used in the RCTs were 2D web-based (Ke, 2008, Segers and Verhoeven, 2005; Baños *et al.*, 2013; Wang, 2008) which were single-player and non-collaborative. There was one collaborative 3D Virtual World in a museum (Wrzesien and Alcaniz Raya, 2010), One 2D and one 2D/3D Massively Multiplayer Online Game (MMOGs) which are collaborative (Harris, 2008; Suh *et al.*, 2010), one 3D RTS game (Chaung and Chen, 2009) and a 2D Gameboy game (Shin *et al.*, 2011). The majority of the games in the RCT studies were 2D games and single-player. Ke (2008) looked at the differences in learning dynamics (individualistic, cooperative and competitive) in terms of looking at collaboration. There was also a collaborative virtual world and two MMOGs

indicating that more study is required with regards to the effectiveness of collaborative games in comparison with single player games. Another area of study which presents itself from these results is that there is no study comparing the effectiveness of 2D games in comparison with 3D games in PE.

19 quasi-experimental studies were found in this category. 15 reported positive results however it should be noted that only six of these studies had a control group (Chen *et al.*, 2010; Çoban and Tuncer, 2008; Hung *et al.*, 2009; Korallo *et al.*, 2012; Yang, Chen and Jeng, 2010; Yien, Hung, Hwang and Lin, 2011). 4 of the studies reported a neutral result where two of them utilised a traditional teaching approach (Petros and Georgios, 2011; Vogel *et al.*, 2006). The majority of quasi-experimental studies (12, 63%) used 2D games and six studies (31.5%) used 3D games. One study used a 2D and a 3D game. Harrington (2009) presented empirical evidence to advocate the use of a 3D game as leading to higher knowledge gain as a result of higher navigational freedom and visual fidelity. The majority of quasi-experimental studies were non-collaborative (59%) where one study in particular looked at collaborative and single player play with the result indicating that there was no significant difference with regards to learning. 8 of the quasi-experimental (42%) studies were comparing the experimental game groups to a traditional teaching approach.

1 quantitative survey used a large amount of participants and multiple regression analysis (Kim and Chang, 2010) where male language minority students who played Math computer games demonstrated higher Math performance scores compared with their male English-speaking counterparts who did not play.

### *3.6.2.2 Perceptual and cognitive skills*

6 studies were identified in this category. 2 studies were RCTs where one study looked at self-perception in relation to mental computational skills, obtained a positive result and used a traditional control group (Miller and Robertson, 2011) and one study looked at intrinsic motivation, flow, cognitive load and learning in relation to background music which obtained a neutral result and did not have control group (Linek and Marte, 2011). Both RCTs utilised 2D games. The majority of the studies in this category (4, 66.7%) adopted a quasi-experimental design and provided positive results (Bottino, Ott, and Benigno, 2009; Cha, Baek and Xu, 2008; Ketamo, 2009; Robertson, 2012). Only one quasi-experimental study adopted a traditional group (Ketamo, 2009) and one study was longitudinal (Bottino, Ott and Benigno, 2009)

### *3.6.2.3 Affective and motivational outcomes*

6 high quality studies were found in this category. The majority (4 of 6) were RCTs (Habgood and Ainsworth, 2011; Kuo, 2007; Owston *et al.*, 2009; Wang, Tsai, Chou and Hung, 2010) and 2 were quasi-experimental (Inal and Cagiltay, 2007; Vos, van der Meijden and Denessen, 2011). 3 studies

(50%) looked at 3D games (Habgood and Ainsworth, 2011; Kuo, 2007) but one of these studies used a variety of 3D games for entertainment purposes specifically looking at flow experience (Inal and Cagiltay, 2007) were the others used games for learning. The three studies looking at 3D games were all single player games and did not use a traditional teaching approach. Two of the studies were for Mathematics and Science (Habgood and Ainsworth, 2011; Kuo, 2007), however, the study looking at flow theory did not attempt to teach any subject. Three studies (50%) used 2D games and also employed traditional teaching approaches, however, two of these studies specifically focussed on games-based construction learning for the purposes of learning Computer Science (Vos, van der Meijden and Denessen, 2011) and Literacy (Owston *et al.*, 2009). One 2D study used GBL for the purposes of teaching Maths and Languages (Wang, Tsai, Chou and Hung, 2010). All these studies had positive results and were single player games with the exception of one study not specifying player mode (Owston *et al.*, 2009)

#### *3.6.2.4 Behavioural change*

4 high quality relevant studies were found in this category. The majority were quasi-experimental (Dias and Agante, 2011; Sun, Wang and Chan, 2011; Tsai, 2008) indicating insufficient numbers, no randomisation or no control group utilised in the study. One study employed an RCT and utilised a traditional teaching group (Vannini *et al.*, 2009). All studies in this category were positive and studied single player games. Two studies used a 3D game (Tsai, 2008; Vannini *et al.*, 2009) and two a 2D game (Dias and Agante, 2011; Sun, Wang and Chan, 2011). The majority of the studies in this category were in the field of health and wellbeing (Dias and Agante, 2011; Tsai, 2008; Vannini *et al.*, 2009) specifically nutrition and empathy.

## **4. Discussion**

This study is the result of searching 17 electronic databases and a total of 18,298 papers. The findings were 105 papers that were considered relevant for inclusion and, after the multi-dimensional framework criteria were applied, 45 were considered to be high quality scoring 9 or above out of 15. The multi-dimensional framework was considered to be useful in providing an exceptionally thorough, rigorous method of categorising papers, not only for the purpose of scoring but also for extracting intricate information about individual studies detailed in Appendices A – D. Overall the framework categorises papers using five pieces of criteria: appropriateness of research design for answering the research questions and sub-questions, appropriateness of methods and analysis, generalisability of the findings, relevance of the particular focus of the study for answering the research questions and the sub-questions and the trustworthiness of the study findings in answering the research questions. Other pieces of vital information were also extracted such as: the research methodology, the educational field of application e.g. Maths, Science, whether the game used was

collaborative, whether a traditional teaching approach was used for the control group and whether the game was 2D or 3D. This assisted us in selecting the studies and compiling the descriptions in the appendices.

The majority of the 105 papers were quantitative (74.3%). The most popular research designs from the quantitative studies were quasi-experimental (55.2%) followed by RCT (17.1%) and then survey (1.9%). 27 studies reporting qualitative data primarily performed case studies, observational studies during the game play or mixed methods where no statistically significant results could be recorded. In comparison with a previous similar study (Connolly *et al.*, 2012) there was no strong preponderance for the use of the survey methodology suggesting that perhaps surveys are not as suitable for conducting studies at PE level than at Secondary Education and Higher Education levels.

In terms of the 45 high quality papers discovered, 40 studies were in the games for learning category of which 23 adopted a quasi-experimental design, 16 were RCTs and one was a survey. There were five studies that used a game for entertainment and all of these studies adopted a quasi-experimental design. The majority of studies in PE used a specially implemented game for learning.

Game genres in the studies were very varied. Strategy games were the most popular genre followed by puzzle games, simulation games, role-playing games, adventure games and generic games. When specifically focusing on entertainment games used for learning the most popular genre was simulation, followed by strategy then generic, puzzle and adventure. With regards to games for learning the most popular genre was strategy, followed by puzzle, role-playing, adventure and simulation. The least popular genres associated with games for learning were action, animated tutorial, mobile, virtual reality, virtual world, GBCL and sports. None of the 105 shortlisted papers used fighting games, platform games or racing games. This indicates that there are certain types of genres that are more suitable for PE than others and also genres that educationalists can specifically tailor for their learning outcomes particularly strategy, puzzle and simulation games.

The most popular delivery platform overall was the PC, followed by the video game console, online games and mobile games. When specifically looking at entertainment games used in the studies it was the video game console followed by the PC. In terms of customised games for learning, the PC was the most popular, then the video game console and mobile. In the area of PE the results indicate that the best method of delivery for educational games is the PC.

In terms of the subject disciplines that the games in the studies were applied to, the majority of the games were in the areas of mathematics, science, languages and social areas which is not really particularly surprising in PE as there is a strong emphasis placed on these subjects at this level of education. The least popular subjects in terms of GBL application were personal development, music, general knowledge and cross curricular games.

One of the purposes of this systematic literature review was to analyse the outcomes and impacts of playing games. As expected the highest quantity of studies associated with PE were in the knowledge acquisition and content understanding category, followed by affective and motivational outcomes, perceptual and cognitive outcomes, behavioural change and social/soft skill outcomes respectively. The most frequently studied effects outcome with games for learning was knowledge acquisition and content understanding, followed by perceptual and affective and motivational skills and perceptual and cognitive skills. Entertainment games that were used for the purposes of learning generally followed the same trend, however, slightly more entertainment games were used for the purposes of learning that investigated behavioural change.

Quasi-experimental designs were the most frequently used methodology when studying knowledge acquisition, affective and motivational skills, perceptual and cognitive skills, behavioural change and social/soft skill outcomes although a few studies on knowledge acquisition also used RCTs, qualitative designs and surveys.

For the purposes of presenting a synthesis of the information found in the systematic literature review it was decided to look at experimental design, collaborative game play, use of a traditional teaching control group, subject, type of game (2D or 3D) and whether the results of the study were positive. The purpose of this was to assess whether there is sufficient empirical evidence for the use of GBL at PE level.

This study has attempted to answer the research question of “what empirical evidence is there concerning the positive impact and outcomes of computer games at PE level?” Taking all results into account indicate that more RCTs are required, a greater amount of comparisons are required using traditional teaching approaches, a greater amount of studies are required examining collaborative gameplay and a greater amount of studies are required investigating the pedagogical benefits of using 2D or 3D games. Moreover, a wider variety of subject areas could be explored and this study has highlighted a number of areas where GBL has not been applied particularly frequently such as personal development, music and general knowledge. Longitudinal studies are extremely scarce in the digital games for learning literature and more are required to appropriately investigate the long term effects of computer games on learning (Brom, Sisler and Slavik, 2010). Goldstein (1968) describes a longitudinal study as one where “*information is repeatedly collected, overtime, on the same sample of individuals.*” Longitudinal studies are incredibly useful for looking at changes over time (possibly spanning decades) and are useful for medical research and lifespan issues.

A limitation of this study which is also a concern in Connolly *et al.*, (2012) was the emphasis on RCT and quantitative studies. This brings into question the rating of study quality where RCT methods are

high quality, quasi-experimental are medium and single subject studies and qualitative studies are low. The emphasis on RCTs may be important in Psychology studies, however in educational studies RCTs may not be the best practice. It has been suggested that the rating system should be seriously reconsidered in future studies and one possible step would be to look at peer reviewed studies which are high quality regardless of experimental design.

## 5. Conclusion

This paper has presented a systematic literature review identifying high quality empirical studies in the area of GBL in PE over a 13 year period from 2000 to the middle of 2013. The review adopted a multi-dimensional framework developed by Connolly *et al.* (2012) to rigorously identify highest quality empirical papers. Initially 105 papers were identified, but after applying the multi-dimensional framework this reduced to 45 high quality empirical papers. This paper has particularly focused on study design, whether the game used was digital or non-digital, the game genre, delivery platform, subject and curricular area along with learning outcomes and impacts. The learning outcomes and impact categories identified were knowledge acquisition and content understanding, affective and motivational outcomes, perceptual and cognitive impacts and behavioural change. The greatest quantity of papers (29) were found in the knowledge acquisition and content understanding category. Six papers were found in the affective and motivational category, six were found in the perceptual and cognitive category and four were found in the behavioural change category.

This study has provided a valuable repository of empirical evidence for researchers and educationalists and is a snapshot of the state of the empirical evidence in this period (2000-2013). The following future research directions became evident:

- Further RCTs should be performed where possible comparing GBL to traditional teaching approaches to ascertain if GBL are a viable teaching approach in PE level. Quasi-experimental designs can also compare GBL to traditional approaches and could be beneficial where RCTs are not practically possible depending on the research circumstances..
- Further longitudinal studies are required as only one was discovered in this systematic literature review.
- Further studies are required looking at whether there are pedagogical benefits of using 2D or 3D games at PE level to identify benefits or disadvantages of using 2D or 3D games. 3D games are reputed to be more immersive than 2D games and it would be beneficial to see if advanced graphics and sound are required for an effective learning experience.
- Further studies are also required to perform comparisons between single and collaborative play to identify pedagogical benefits.

There has been an extensive amount of data collected associated with the use of GBL in PE and as a result we have the information required to perform a number of quantitative meta-analyses to generate meaningful statistical data. As one of the limitations of this study is that it only ranges from 2000 to the first half of 2013 then another future research direction will be to attempt to fill in the gap to the present

## References

- Abt, C. C. (1987). *Serious Games*. University Press of America.
- Annetta, L., Mangrum, J., Holmes, S., Collazo, K. & Cheng, M.-T. (2009). Bridging Reality to Virtual Reality: Investigating gender effect and student engagement on learning through video game play in an elementary school classroom. *International Journal of Science Education*, 31, 1091-1113.
- Baños, R.M., Cebolla, A., Oliver, E., Alcañiz, M. & Botella, C. (2013). Efficacy and acceptability of an internet platform to improve the learning of nutritional knowledge in children: the ETIOBE mates. *Health Education Research* 28 (2), pp 234–48.
- Bottino, R.M., Ott, M., & Benigno, V. (2009). Digital mind games: Experience-based reflections on design and interface features supporting the development of reasoning skills. In *Proceedings of the 3<sup>rd</sup> European Conference on Games-based Learning (ECGBL)*, 12 - 13 October 2009, Graz, Austria.
- Boyle, E., Hainey, T., Connolly, T.M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Pereira, J., & Riberio, C. (2015), "An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games." *Computers & Education*, Vol. 94, No. 2, pp. 178-192.
- Brom, C., Sisler, V. & Slavik, R. (2010). Implementing Digital Game-Based Learning in Schools: Augmented Learning Environment of 'Europe 2045'. *Multimedia Systems*, Volume 16, Number 1, 23-41.
- Cha, J., Baek, Y. & Xu, Y. (2008). Exploring Learner's Variables Affecting Gaming Achievement in Digital Game-Based Learning. 75-82.
- Chatterjee, S., Mohanty, A. & Bhattacharya, B. (2011). Peer Collaboration, Facilitator Intervention, and Learning Styles in Computer Game-Based Learning: Initial Findings From and Empirical Study. In *Proceedings of the 5<sup>th</sup> European Conference on Games-Based Learning*, 20-21 October, Athens, Greece.
- Chen, H.-P., Lien, C.-J., Annetta, L. & Lu, Y.-L. (2010). The Influence of an Educational Computer Game on Children's Cultural Identities. *Educational Technology and Society*, 13 (1), 94–105.
- Chen, S., & Michael, D. (2005). *Serious Games: Games that Educate, Train and Inform*. USA, Thomson Course Technology
- Cheng, H. N. H., Wu, W. M. C., Liao, C. Y. & Chan, T-W. (2009). Equal Opportunity Tactic: Redesigning and Applying Competition Games in Classrooms. *Computers and Education*, 53, 866 – 876.
- Chuang, T.-Y. & Chen, W.-F. (2009). Effect of Computer-Based Video Games on Children: An Experimental Study. *Educational Technology and Society*, 12, 1-10.
- Çoban, S. & Tuncer, İ. (2008). An Experimental Study of Game-Based Music Education of Primary School Children. In *Proceedings of the 2<sup>nd</sup> European Conference on Games-Based Learning*, 16-17 October, Barcelona, Spain.
- Cobb, T., & Horst, M. (2011). Does Word Coach Coach Words? *CALICO Journal*, 28(3), p-p 639-661.
- Connolly, T. C., Boyle, E. A., Hainey, T. Macarthur, E. & Boyle, J. M. (2012). A Systematic Literature Review of Empirical Evidence on Computer Games and Serious Games. *Computers and Education*, 59, 661 – 686.

De Freitas, S. (2006). Learning in immersive worlds, Bristol Joint Information Systems Committee, Retrieved May 1, 2016, from:  
[http://www.jisc.ac.uk/media/documents/programmes/elearninginnovation/gamingreport\\_v3.pdf](http://www.jisc.ac.uk/media/documents/programmes/elearninginnovation/gamingreport_v3.pdf)

- Dempsey, J. V., Rasmussen, K. & Lucassen, B. (1994). Instructional gaming: implications for instructional technology. Paper presented at the *Annual Meeting of the Association for Educational Communications and Technology*, 16-20 February 1994, Nashville, TN.
- Dias, M., & Agante, L., (2011). Can advergames boost children's healthier eating habits? A comparison between healthy and non-healthy food. *Journal of Consumer Behaviour*. 10(3), 152 – 160.
- Egenfeldt-Nielsen, S. and T. Buch (2006). The learning effect of 'Global Conflicts: Middle East'. In Gaming Realities: A Challenge for Digital Culture. Athens: 93-97.
- Egenfeldt-Nielsen, S., Smith, J. H., & Tosca, S. P. (2008). *Understanding Video Games: The Essential Introduction*. Routledge.
- Furió, D., González-Gancedo, S., Juan, M. C., Segui, I. & Rando, N. (2013). Evaluation of learning outcomes using an educational iPhone game vs. traditional game. *Computers and Education*, 64, 1 – 23.
- Garzotto, F. (2007). Was Vygotsky right? evaluating learning effects of social interaction in children internet games. *Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction - Volume Part II*. Rio de Janeiro, Brazil: Springer-Verlag.
- Goldstein, H. (1968), "Longitudinal Studies and the Measure of Change. Journal of the Royal Statistical Society." Vol. 18 No. 2, pp. 93-117.
- Griffiths, M.D. (2002). The educational benefits of videogames. *Education and Health*. Vol. 20, Issue. 3.
- Hainey, T. (2010). *Using Games-Based Learning to Teach Requirements Collection and Analysis at Tertiary Education Level*. Doctor of Philosophy, University of the West of Scotland.
- Hainey, T., Connolly, T., Boyle, L. & Azadegan, A. (2014). A Systematic Literature Review to Identify Empirical Evidence on the Use of Games-based Learning in Primary Education for Knowledge Acquisition and Content Understanding. 8th European Conference on Games-based Learning (ECGBL), 9-10 October, Berlin, Germany.
- Habgood, M. P. J. & Ainsworth, S. E. (2011). Motivating Children to Learn Effectively: Exploring the Value of Intrinsic Integration in Educational Games. *Journal of the Learning Sciences*, 20, 169-206.
- Harrington, M. C. R. (2012). The Virtual Trillium Trail and the Empirical Effects of Freedom and Fidelity on Discovery-Based Learning. *Virtual Reality*, 16, 105 – 120.
- Harris, D. (2008). *A comparative study of the effect of collaborative problem-solving in a Massively Multiplayer Online Game (MMOG) on individual achievement*. Doctor of Education, The University of San Francisco.
- Hung, C.-M., Chiu, C.-H., Chen, Y.-T., SU, M.-J. & Chen, H.-S. (2009). Effectiveness of game-based learning of a national health e-learning network for nutrition education in elementary school. *Proceedings of the 11th international conference on e-Health networking, applications and services*. Sydney, Australia: IEEE Press.
- Inal, Y. & Cagiltay, K. (2007). Flow experiences of children in an interactive social game environment. *British Journal of Educational Technology*, 38(3), 455–464.
- Ke, F. (2008). Computer games application within alternative classroom goal structures: cognitive, metacognitive, and affective evaluation. *Educational Technology Research and Development*, 56, 539-556.
- Ketamo, H. (2009). Semantic networks -based teachable agents in an educational game. *WSEAS Transactions on Computers*, 4 (8).
- Kim, P., Buckner, E., Kim, H., Makany, T., Taleja, N. & Parikh, V. (2012). A comparative analysis of a game-based mobile learning model in low-socioeconomic communities of India. *International Journal of Educational Development*, 32, 329 – 340.

- Kim, S. & Chang, M. (2010). Computer Games for the Math Achievement of Diverse Students. *Educational Technology and Society*, 13 (3), 224–232.
- Kuo, M.-J. (2007). How does an online game based learning environment promote students's intrinsic motivation for learning natural science and how does it affect their learning outcomes ? . *The First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL'07)*.
- Korallo, L., Foreman, N., Boyd-Davis, S., Moar, M. & Coulson, M. (2012). Do Challenge, Task Experience or Computer Familiarity Influence the Learning of Historical Chronology from Virtual Environments in 8-9 Year Old Children? *Computers and Education*, 58, 1106 – 1116.
- Linek, S. B., Marte, B. & Albert, D. (2013). *Background Music in Educational Games: Motivational Appeal and Cognitive Impact*. In Developments in Current Game-Based Learning Design and Deployment (Ed: P. Felicia), Premier Reference Source.
- Meluso, A., Zheng, M., Spires, H. A. & Lester, J. (2012). Enhancing 5th graders' science content knowledge and self-efficacy through game-based learning. *Computers and Education*, 59, 497 – 504.
- Miller, D. J., & Robertson, D. P. (2011). Educational benefits from using games consoles in a primary classroom: A randomised controlled trial. *British Journal of Educational Technology*, 42 (5), 850 – 864.
- Oberschain, F. B. (2010). Using handheld computers for constant daily review in sixth grade mathematics. *Doctoral Dissertation*, College of Education and Humane Development. Retrieved 20 December 2012 from: <http://hdl.handle.net/1920/5796>.
- Owston, R., Wideman, H., Ronda, N. S. & Brown, C. (2009). Computer games developed as a literacy activity. *Computers and Education*, 53 (3), 977 – 989.
- Petros, P. & Georgious, F. (2011). Educational Applications of Serious Games: The Case of the Game Food Force in Primary Education Students. *In Proceedings of the 5<sup>th</sup> European Conference on Games-Based Learning (EC-GBL)*, Athens, Greece.
- Robertson, J. (2012). Making games in the classroom: Benefits and gender concerns. *Computers and Education*, 59, 385-398.
- Rubin-Vaughan, A., Pepler, D., Brown, S., & Craig, W. (2011). Quest for the Golden Rule: An effective social skills promotion and bullying prevention program. *Computers and Education*, 56(1), 166–175.
- Segers, E. & Verhoeven, L. (2005). Long-Term Effects of Computer Training of Phonological Awareness in Kindergarten. *Journal of Computer Assisted Learning*, 21(1), 17-27.
- Shin, N., Sutherland, L. M., Norris, C. A. & Soloway, E. (2011) Effects of game technology on elementary student learning in mathematics. *British Journal of Educational Technology (BJET)*, 43 (4), 540 – 560.
- Suh, S., Kim, S.W. & Kim, N. J. (2010). Effectiveness of MMORPG-based instruction in elementary English education in Korea. *Journal of Computer Assisted Learning*, 26, 370–378.
- Sun, C-T., Wang, D-Y. & Chan, H-L. (2011). How digital scaffolds in games direct problem-solving behaviors. *Computers and Education*, 57(3), 2118 - 2125.
- Tang, S., Hanneghan, M. & El Rhalibi, A. (2009). Introduction to Games-Based Learning, In Games-based Learning Advancement for Multisensory Human Computer Interfaces: Techniques and Effective Practices (Eds: T.M. Connolly, M.H. Stansfield and E. Boyle). Idea-Group Publishing: Hershey. ISBN: 978-1-60566-360-9
- Tsai, Y-F. (2008). *The effects of interacting with a computer-simulated virtual pet dog on children's empathy and humane attitudes*. Doctoral Dissertation, Retrieved 30 August from: <http://summit.sfu.ca/item/9146>
- Vannini, N., Enz, S., Sapouna, M., Wolke, D., Watson, S., Woods, S., Dautenhahn, K., Hall, L., Paiva, A., Andre, E., Aylett, R., & Schneider, W. (2010). FearNot!: a computer-based anti-bullying-programme designed to foster peer intervention. *European Journal of Psychology of Education* , vol 26 , no. 1 , pp. 21-44 .
- Vogel, J. J., Greenwood-Ericksen, A., Cannon-Bowers, J. & Bowers, C. A. (2006). Using Virtual Reality with and without Gaming Attributes for Academic Achievement. *Journal of Research on Technology in Education*.

- Vos, N., van Der Meijden, H. & Denessen, E. (2011). Effects of constructing versus playing an educational game on student motivation and deep learning strategy use. *Computers and Education*, Vol. 56, No. 1, 127-13.
- Wang, T. (2008). Web-based quiz-game-like formative assessment: Development and evaluation. *Computers and Education*, 51(3), 1247–1263.
- Wang, H. S., Tsai, S. N., Chou, C. H. & Hung, H. J. (2010). The study of motivation and reasoning faculties of game-based learning in elementary school students. In *Education Technology and Computer (ICETC), 2010 2nd International Conference on* (Vol. 3, pp. V3-247). IEEE.
- Wilson, A., Hainey, T. & Connolly, T. M. (2013). Using Scratch with Primary School Children: An Evaluation of Games Constructed to Gauge Understanding of Programming Concepts. *International Journal of Games-Based Learning*. Vol. 3, Issue 1.
- Woolfson, L. M. (2011). *Educational psychology: The impact of psychological research on education*. London: Prentice Hall, Pearson Education.
- Wrzesien, M. & Alcaniz Raya, M. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project. *Computers and Education*, 55(1), 178–187.
- Yang, J. C., Chen, C. H. & Jeng, M. C. (2010). Integrating video-capture virtual reality technology into a physically interactive learning environment for English learning. *Computers and Education*, 55, 1346 – 1356.
- Yien, J-M., Hung, C-M., Hwang, G-J. & Lin, Y-C. (2011). A Games-Based Learning Approach to Improving Students' Learning Achievements in a Nutrition Course. In *Proceedings of the Turkish Online Journal of Educational Technology*, 10(2), 1-10.
- Zyda, M. (2005). From Visual Simulation to Virtual Reality to Games. *Computer*, 38, 25-32.

### Appendix A: Knowledge acquisition/Content understanding outcomes

Author(s)	Aims/Objectives of Study	Methods	Conclusions
Annetta et al. (2009)	Examining learning of simple machines and engagement using Dr. Friction.	Mixed-method pre-test/post-test design utilised. 74 5 <sup>th</sup> -grade students aged 10-11 participated (31M and 43F) played Dr. Friction for several days mid-unit. ANCOVA was utilised.	No significant difference for gender (0.133), $f=2.33$ in relation to knowledge gain. Overall gain from pre-test to post-test was significant (0.000), $f=67.02$ . Observations/focus groups suggested high engagement and video game technology can scaffold learning.
Baños et al. (2012)	Studying efficacy and acceptability of online game called 'ETIOBE Mates' to improve nutritional knowledge.	Sample of 228 children (10-13 years, mean age=11.22, SD=0.92) from 4 Spanish schools of similar socioeconomic status. Schools were randomly divided into: an experimental group who used game (N=73; 37%M; age=11.2) and a control group given a pamphlet with nutritional information (N=155; 43%M; age=11.2). ANOVAS and chi squared tests were utilised.	Both groups increased scores for nutritional knowledge but the game group acquired more knowledge. Results suggested that the younger group liked the game more.
Chatterjee et al. (2011)	Investigating the effect of different pedagogical contexts and learning styles on learning outcomes when playing educational computer games.	Involved 231 participants (F = 83, M = 148, Age average = 12.6) from classes 7 and 8 of five East Indian schools. All participants had English as the medium of school instruction. Adopted a mixed-method, quasi-experimental. Based on <i>Global Conflicts-Sweatshops</i> and <i>Playing History-The Plague</i> . Participants were assigned to one of 4 groups: Collaborative with facilitator; Collaborative without facilitator; Individualistic with facilitator; Individualistic without facilitator intervention. First 2 groups each had 59 participants, while other two had 58 and 55 respectively. All participants played the two games. ANOVA used to explore differences in learning outcomes.	Students benefit more from educational computer games in a pedagogical context comprising both peer collaboration and facilitator intervention.
Chen et al. (2010)	Exploring influences an educational computer game has on cultural identities.	Pre-test/post-test experimental/control design was carried. 64 (35M, 29F) in the experimental group and 66 (36M, 30F) in control group. Experimental group used the game as a supplemental activity and the control group used normal instruction. ANCOVA compared scores in terms of cultural identities of the groups. Paired t-tests were utilised.	Experimental group significantly strengthened their cultural identities. The game had a significant effect on cultural learning and was helpful to students from lower socioeconomic families.

Cheng et al. (2009)	Evaluating a computerised equal opportunity tactic to match students who have similar ability in competition (EOT) using a competitive learning game, AnswerMatching.	<p>3 groups of 3<sup>rd</sup> year students were assigned to an EOT (similar performance paired) group (N=24), RAN (paired randomly) group (N=30) and an HTL (high-to-low - more-able paired with a less-able) group (N=26) based on performance.</p> <p><b>RQ1:</b> A 3 (EOT, RAN, and HTL) x 2 (more-able and less-able students) ANOVA on overall scores was adopted.</p> <p><b>RQ2:</b> A 3x2 ANCOVA was carried out, with group and ability as the between-subject variables.</p>	<p><b>RQ1:</b> EOT balanced the perceived performance in a classroom without changing its mean.</p> <p><b>RQ2:</b> When either EOT or RAN was adopted in the game, the perceived performance of students was not influenced by their abilities eventually.</p>
Chuang and Chen (2009)	Comparing traditional and games-based instruction for fire safety education using Fire Department 2: Fire Captain.	Pre-test/post-test experimental/control group design of 3 <sup>rd</sup> -grade students (61M and 54F participants). 58 participants (control) 57 (experimental). Utilised MANOVA and univariate analysis.	Results showed a significant difference between computer-assisted instruction and video game playing in learning achievement. MANOVA showed significantly higher knowledge scores in treatment groups ( $F=2.739$ , $p<0.047$ ).
Çoban and Tuncer (2008)	Investigating teaching "Rhythm in Music" using the games "Bingo", "Wheel".	A pre-test/post-test methodology with an experimental (game) and control group (rhythm patterns written on board). Experiment involved 6 <sup>th</sup> grade students aged 12-13 (N=52) with 26 students in each group (44%F, 56%M). T-tests were used.	After four weeks the skills and capabilities differed substantially according to the teaching method used in the classroom ( $t=7.422$ , $df=25$ $p<0.01$ ).
Cobb and Horst (2011)	Investigating an integrated suite of vocabulary training games called <i>My Word Coach</i> .	Study used a quasi-experimental, within-subjects design where one group used game for 2 months while other served as quasi-control, and then roles were reversed. Participants were 2 full classes of 25 grade six Francophone ESL learners in a middle-class suburban school in Montreal (11-12 years old). Ethnicity of both groups was roughly 30% Quebec Francophone children and 70% immigrant children from mainly Francophone countries. Results analysed using one-way ANOVAs followed by Tukey post-hoc tests.	Results suggested that <i>My Word Coach</i> helped many of the learners develop their English lexicons on both declarative and procedural levels. Results indicated an average vocabulary expansion of between roughly 10% - 20% in a 4-month period.
Furió et al. (2013)	Determining if a mobile game with augmented reality (AR) helped children in the learning process.	Repeated measures experimental control group after participants played the game with N=84 (38M, 46F), aged 8-10. Students randomly assigned to one of two conditions: a) those who played the AR game first and then the traditional game; b) those who played the traditional game first and then the AR game. ANOVA, MANOVA, t-tests used to	The results indicate that the students achieved similar knowledge improvements using an AR game as when using a custom and guided game (traditional game).

		analyse the results.	
Garzotto (2007)	Investigating educational effectiveness, content, enjoyment and social interaction of a multiplayer game called Pirates Treasure Hunt.	Pre-test/post-test experimental design utilised with 85 participants from a public elementary school in Milan. 44F and 41M participants; 65 participants were 7-8 years old (from 3 2 <sup>nd</sup> -grade classes), and 20 were aged 9-10 (from a 4 <sup>th</sup> -grade class). Learning effectiveness was analysed using a paired samples <i>t</i> -test.	The correct answers in the post-test were significantly greater than in pre-test ( $t(84)=10.25$ , $p<0.01$ ), mean pre-test=5.14; mean post-test=7.16).
Harrington (2009)	Investigate graphical fidelity and navigational freedom on learning outcomes in educational.	Volunteer sample of 64 3 <sup>rd</sup> -5 <sup>th</sup> grade students engaged in the experiments. All of the volunteers were interested in computers, video games, nature, and art. Students randomly assigned to one of 4 groups: (1) High Navigational Freedom (NF) and Low Visual Freedom; (2) High NF and High VF; (3) Low NF and Low VF; (4) Low NF and High VF. Pre-test and post-test used. Two teachers and the researcher graded the pre/post-tests and logs.	-High VF ( $M=30.95$ , $SD=14.76$ ) produced higher scores for Knowledge Gain than Low VF ( $M=19.99$ , $SD=13.39$ ) ( $F(1,60)=10.54$ , $p=0.0019$ ). -High NF ( $M=28.24$ , $SD=16.51$ ) produced slightly higher Knowledge Gain scores than Low NF ( $M=22.69$ , $SD=13.06$ ). -Two-way ANOVA produced significant evidence of interaction ( $F(1,60)=4.85$ , $p=0.0315$ ) of VF and VF on Knowledge Gain. -Most powerful learning is achieved combining High VF and High NF.
Harris (2008)	Comparing collaborative problem-solving in a MMOG (Web Earth Online) on individual achievement.	159 6 <sup>th</sup> -grade students from a junior-high school in San Jose, California. 4 comparison classes ( $n=100$ ), 2 taught by the researcher and 2 by the regular science teacher. There was an experimental group of two existing classes ( $n=59$ ) taught by researcher. There were almost a third as many females as males. 65 participants were 12, 93 were 11 and one was 13 years old. A dependent-samples <i>t</i> -test was performed.	The MMOG group did not have greater improvements in scores on content knowledge unit.
Hung et al. (2009)	Studying the effects of using games to teach nutrition.	33 3 <sup>rd</sup> -grade students split into experimental (game) and control groups (lectures and multimedia slide show) with pre-test/post-test.	Experimental group received better scores on the nutrition knowledge test and dietary behaviour assessments. Very little difference between the two groups in nutrition attitude. Gender analysis revealed no significant difference in the experimental group.
Ke (2008)	Examining whether alternative classroom goal structures would enhance the effects of computer games using ASTRA EAGLE for Maths.	Pre-test/post-test quasi-experimental design with 358 students recruited from eighteen 5 <sup>th</sup> -grade public school classes in 4 rural Pennsylvanian school districts. Participants varied in gender, socio-economic status and prior maths ability level: 49%	Computer games were significantly more effective in promoting learning motivation but not in facilitating cognitive maths test performance and metacognitive awareness. Cooperative goal structure significantly enhanced the effects of computer games on attitudes toward math learning.

		female, 38% economically disadvantaged; for prior maths ability, 23% were below basic, 20% basic, 34% proficient and 23% advanced.	
Kim et al. (2012)	Exploring effectiveness of a game-based mobile learning model for children living in underdeveloped regions.	210 children aged 6-14 ( $M=11.12$ , $SD=1.24$ ) from 6 locations in India. Two community settings were selected: urban slums and rural villages. Children were randomly assigned into one of 3 groups: one child per device (G1); three per device (G3) and seven per device (G7). All participants were equipped with TeacherMate, a handheld mobile learning device, with a Fire Rescue Math (FRM) educational game. A 3 (Group formation) x 6 (Location) x 3 (Gender = Boy, Girl, Mixed) univariate ANOVA was used.	Children with little or no previous exposure to technology were able to figure out the given mobile learning technology and solve a series of incrementally challenging problems without specific adult intervention. Gender and group size, do affect ability to adopt and learn.
Kim and Chang (2010)	Examining effects of playing computer games on maths achievement.	Study used the 2005 National Assessment of Educational Progress (NAEP). NAEP gathers data every year on US 4 <sup>th</sup> , 8 <sup>th</sup> and 12 <sup>th</sup> -grade student achievement scores in subjects such as reading, maths, writing, science, US history and civics. Study performed regression analyses using more than 170,000 U.S. 4 <sup>th</sup> -grade maths results.	English-speaking students who played computer maths games in school every day displayed significantly lower maths achievement than those who never played. Male language minority students who daily played computer games in maths demonstrated higher maths performance scores compared with their male English-speaking counterparts who never played.
Korallo et al. (2012)	Investigating whether challenge, task experience or computer familiarity influence the learning of historical chronology from virtual environments in 8–9 year old children.	<b>Experiment 1:</b> 52 children (32M, 20F) drawn from a single class at a primary school in North-East London (average age 8.5). Children were divided into 3 groups on a pseudorandom basis: Paper group (N=16; 8M, 8F), PowerPoint group (N=18; 12M, 6F), and a Virtual Environment (VE) group (N=20; 14M, 6F). <b>Experiment 2:</b> 45 primary school children (aged 8–9) were allocated to 3 independent groups on same basis as in Experiment 1, trained using Paper (N=15; 11M, 4F), a VE (N=15; 10M, 5F) and PowerPoint (N=15; 10M, 5F). <b>Experiment 3:</b> 30 pupils (14M, 16F) aged 8–9 from Ukraine.	No significant effect was found between the 3 groups (with minimal pre-training). Error free learning occurred when greater exploration was permitted prior to training in the VE. In Ukraine, children produced better learning compared to PowerPoint, but no better than in a Paper condition. Results confirmed the benefit of using challenge in a VE, but only with adequate prior medium familiarisation.
Meluso et al. (2012)	Investigating effects of collaborative and single player conditions on science content	100 5 <sup>th</sup> -grade students (45M and 55F), approx. 19% were African American, 7% Asian, 44% Caucasian, 18% Latino and 12% other. Across 6 classes,	Results indicated no differences between the two playing conditions; however, overall, science content learning and self-efficacy increased.

	learning and self-efficacy.	students were randomly assigned to either a single-player or a collaborative gameplay condition. Students played the Crystal Island game after exposure to curriculum landforms that related to North Carolina standard course of study for 5 <sup>th</sup> grade science.	
Obenschain (2010)	Examining Maths achievement by ability and gender after using handheld computer games or teacher-led reviews.	54 6 <sup>th</sup> -grade students participated in the study using a within-group equivalent time series design. Pre-test/post-test design with multiple measures was used to measure students' retention of Maths concepts for low-ability, average-ability, and high-ability learners and for male and female learners. A delayed post-test was used to analyse retention later. Qualitatively, the teacher kept a journal.	Experimental game group had a mean gain score on units 1, 3 and 5 tests of 82.28 ( $SD=30.00$ ) and students who reviewed using the teacher-led reviews had a mean gain score on units 2, 4 and 6 tests of 58.81 ( $SD= 45.64$ ). Results from the paired samples $t$ -test were statistically significant, $t(52)=4.42$ , $p<.001$ . Average-ability students using handheld computer Maths review games retained more Maths content on unit tests. Average-ability female students using handheld computer Maths review games retained more Maths content.
Petros and Georgios (2011)	Investigating the use of Food Force as a learning tool.	34 participants (15M, 19F), aged 10-11; divided into 2 groups: experimental group (Food Force), N=13 students (6M, 7F) and the control group (text and photo viewer) N=21 (9M, 12F).	Game gave no significantly different results in knowledge construction but contributed significantly to attitudes towards feeding problems in different regions of the Earth ( $t=-3.323$ , $df=12$ -tailed $p=0.006<0.01$ ). The game was significantly more interesting and motivating.
Rubin-Vaughan et al. (2011)	Exploring learning with web-based gaming modules dealing with bullying prevention.	307 2 <sup>nd</sup> to 6 <sup>th</sup> grade children completed the Bark Academy module, 226 for Mission to Mars and 448 for Ghoul School. Pre and post-tests were used to measure knowledge and attitudes.	Significant improvements in knowledge and attitudes across each of the 3 modules. Students enjoyed using the games.
Segers and Ludo (2005)	Examining long-term effects of a computer application for the development of phonological awareness between native Dutch and immigrants at kindergarten level.	100 2 <sup>nd</sup> year kindergarten children with an average age of 5.5 years participated. Experimental group (16 native, 26 immigrants) while control group (22 native, 36 immigrants). Experimental group used CD-ROMs while control group used games occasionally.	The experimental group benefited from the presence of computer games involving letters and showed more progress on grapheme knowledge during the second half of the school.
Shin et al. (2011)	Investigating effects of game technology on learning mathematics.	<b>Study 1:</b> 41 2 <sup>nd</sup> Grade students (7-8 years old) participated. Group1: used digital arithmetic games (using a Gameboy). Group 2: used paper based arithmetic games. After 5 weeks, both groups then played digital arithmetic games for 13 weeks. <b>Study 2:</b> Examined learning in relation to game performance, attitudes toward the game, attitudes toward Maths, gender and ethnicity. During a 4-	<b>Study 1:</b> The digital arithmetic games outperformed those who played the paper-based arithmetic game, and students who played more outperformed those who played the game less. <b>Study 2:</b> Technology-based game influenced students' arithmetic learning. Students who had higher scores on the Gameboy game were more likely to achieve high scores on the arithmetic test

		month period, 50 2 <sup>nd</sup> grade students from 3 classes played a digital arithmetic game. Multiple regression was used to determine the relationship between students' arithmetic scores and characteristics.	regardless of gender, ethnicity or attitude toward the game.
Suh, Kim and Kim (2010)	Investigating differences in achievement between MMORPG-based and face-to-face instruction groups in an English learning environment.	Pre-test/post-test experimental/control group design. 220 elementary school students (5 <sup>th</sup> and 6 <sup>th</sup> -grade) with 118 (experimental who played the game) and 108 (control who used multimedia and textbook).	MMORPGs showed higher scores in listening, reading and writing than face-to-face instruction. In the post-test, differences in mean scores between groups were significant in all subcategories of listening, reading and writing but not speaking. Most influential variables in English learning achievement were prior knowledge, motivation, and network speed in MMORPG-based instruction. Findings suggest that MMORPGs can improve English communicative skills.
Wang (2008)	Examining effectiveness a multiple-choice web-based formative assessment quiz game, in facilitating learning and motivation in an eLearning environment.	166 participants (81M, 84F); 6 <sup>th</sup> -grade students; divided into 3 groups: the PPT (paper-and-pencil test) group N=53 (27M, 26F), the N-WBT (normal web-based test) group N=59 (29M, 30F), and the GAMWATA group N=53 (25M, 28F), a pre-test/post-test experimental/control group design was used and the treatment was for 2 weeks (6 classes).	Students in the GAMWATA group had higher knowledge-and comprehension levels and more actively participated in Web-based formative assessment.
Wrzesien and Raya (2010)	Evaluating E-Junior: a serious virtual world for teaching natural science and ecology.	48 participants (20M, 28F); 6 <sup>th</sup> -grade students (10–11 years) were divided into 2 groups, traditional learning by teacher (control) and the virtual learning using virtual teacher (experimental). Pre-test/post-test experimental/control group design was used. Results were analysed using an ANOVA.	Experimental group did not performed better in knowledge gain ( $F(1,46)=0.699$ , $p=0.408$ ). Students preferred the virtual class ( $F(5,220)=2.859$ , $p=0.027$ ) and were significantly more engaged (post hoc, $p=0.031$ ) in the virtual class ( $M=3.68$ , $SD=0.77$ ) than traditional ( $M=3.08$ , $SD=1.01$ ).
Yang, Chen and Jeng (2010)	Evaluating a Physically Interactive Learning Environment (PILE) by integrating video-capture VR technology into a classroom.	60 2 <sup>nd</sup> -grade participants (35M, 25F) divided into 2 groups, control group (PowerPoint) and experimental group (PILE).	A significant difference ( $p=.043<0.05$ ) in the experimental group students' English learning achievement. The system enhanced the students' learning motivation. The teacher believed this system was beneficial in assisting English learning.
Yien, Hung, Hwang and Lin (2011)	Investigating the influence of a GBL approach on nutrition cognition, attitudes and habits via web-based computer games.	66 3 <sup>rd</sup> -grade participants (36M, 30F) divided into 2 groups, the experimental group (GBL) and control group (Multimedia PowerPoint). Study adopted a pre-test/post-test experimental/control group design and lasted four weeks. Results were analysed using an ANCOVA.	Learning achievement of the experimental group (adjusted mean: 17.39) was significantly higher than the control (adjusted mean: 14.64) ( $F=20.01$ , $p<.001$ ). GBL did not enhance the nutrition attitudes. Learning achievement of the experimental group (adjusted mean: 89.28) was better than that of the control group (adjusted mean: 86.05). GBL can effectively enhance habits ( $F=4.17$ , $p=.05$ , $p<0.05$ ).

Vogel et al. (2006)	Identifying whether a VR learning game can increase learning equally in hearing and deaf children.	<p>Pre-test/post-test quasi-experimental unequal control group design.</p> <p>44 children (7-12 years) from a Florida public elementary school participated (25F, 19M); 12 from 2<sup>nd</sup> grade, 13 from 3<sup>rd</sup>, 9 from 4<sup>th</sup> and 10 from 5<sup>th</sup> grade (2 later excluded). Of remainder, 11 were deaf. Students were split randomly and equally into the control or experimental group..</p>	<p>The difference scores in the deaf/hearing-impaired group from the pre-test to post-test were not statistically significant in Maths or language arts.</p> <p>The deaf control group's difference scores did not change significantly. However, there was a significant improvement in the Maths section (<math>p=.023</math>), suggesting that using the VR program with the CAI method improved Maths skills. No significant changes were observed in the experimental group in Maths or language arts. No significant difference when comparing the change scores between groups in the language arts. However, a significant difference in change scores between groups was found in Maths (<math>p=.033</math>), favouring the CAI version of the VR program over the game.</p>
------------------------	--	--	--

### Appendix B: Affective and motivational outcomes

Author(s)	Aims/Objectives of Study	Methods	Conclusions
Habgood and Ainsworth (2011)	Investigating intrinsic integration in educational games in relation to motivation.	<p><b>Study 1</b> examined learning gains of 58 children between 7 years, 1 month, and 8 years, 10 months (30F, 28M) who played either the intrinsic, extrinsic or control variants of Zombie Division for 2 hours, supported by their classroom teacher. Two-way mixed measures ANOVA (pre-test, post-test and delayed test) X (intrinsic, extrinsic and control) was performed.</p> <p><b>Study 2</b> compared time on task for intrinsic and extrinsic variants of the game when 16 children between ages of 9 years, 10 months and 11 years, 2 months (5F, 11M) had free choice of which game to play. Used a single repeated measure design, game, with 2 levels (intrinsic or extrinsic) that reflected the amount of time children played each game. A pair-samples <i>t</i>-test as utilised.</p>	<p>Children learned more from intrinsic version of the game under fixed time limits and spent 7 times longer playing it in free-time.</p> <p><b>Study 1</b>, All groups improved over the tests: intrinsic <math>F(2,47)=24.89</math>, <math>p&lt;0.001</math>, extrinsic, <math>F(2,47)=6.78</math>, <math>p&lt;0.003</math> and control <math>F(2,47)=3.97</math>, <math>p&lt;0.025</math>.</p> <p><b>Study 2</b> Participants spent more than 7 times longer playing the intrinsic version (75.7 min, <math>SD=35.5</math>) than the extrinsic version (10.28 min, <math>SD=10.28</math>) (<math>t(15)=7.38</math>, <math>p&lt;.001</math>)</p>
Inal and Cagiltay (2007)	Examining flow experiences in an interactive social game.	33 participants (21M, 12F), (20 1 <sup>st</sup> -grade, 5 2 <sup>nd</sup> -grade and 8 3 <sup>rd</sup> -grade, age range 7 - 9). Study lasted for 6 weeks. A flow scale was used in a structured face-to-face interview measuring: challenge, goals, feedback, control, frame story and concentration.	Flow experience in game play happened more in males. Challenge and complexity of games had more effect on flow experiences than clear feedback.
Kuo (2007)	Examining how online GBL promotes students' intrinsic motivation for learning natural science.	Mixed research methods were used to examine "Go Go Bugs". Two classes of 3 <sup>rd</sup> grade students (aged 10) in Taiwan were randomly assigned to a control group (using a multimedia learning environment) ((N=19, 10M and 9F) and the other class (N=27, 16M, 11F) to an experimental group (using a GBL environment).	Results showed significant improvement in interest in learning natural science in GBL and successfully motivated participants in exploring natural science and engaging in learning. No significant results showing that the GBL environment improved students' learning achievement.
Owston et al. (2009)	Examining computer game development as a pedagogical activity to motivate and engage.	Pre-test/Post-test, experimental/control design. 311 4 <sup>th</sup> grade participants. 125 experimental group (60F, 65M) and 186 control students (90F, 96M). A repeated measures MANOVA design contrasting both groups and an ANCOVA was utilised.	285 games were created with considerable variation in the number of games produced by different classes. Median number of games created per class was 31, while the most productive class developed 54 games least productive 20 games. The games contained 7,199 questions. The number of questions written per class ranged from 276 - 1,738 with a median of 726.
Vos, van der Meijden and	Exploring motivations of students who play and	235 students from 9 classes split over 5 <sup>th</sup> and 6 <sup>th</sup> grade, 10-12 years. Intrinsic motivations and deep	Findings showed a significant difference between the perceived competence of the students in both

Denessen (2011)	create their own games.	strategy use were measured with 5 classes constructing their own drag and drop memory games (to master Dutch proverbs) and 4 classes played similar style drag and drop games.	conditions, $F(1,229)=8.49$ , $p=.004$ . Students who constructed the game felt more competent than the students who played the game $F(1,228)=73.69$ , $p<.001$ . The large effect size ( $d=1.07$ ) was in favour of the game constructors. Students in the construction condition rated their effort significantly higher than the play condition, $F(1,229)=74.18$ $p<.001$ .
Wang, Tsai, Chou and Hung (2010)	Exploring the effect of GBL on motivation and reasoning ability.	124 6 <sup>th</sup> -grade students (64M, 60F), divided into a control, N=60 (30M, 30F) and an experimental group, N=64 (34M, 30F); participants in the experimental group used a "computer gaming teaching" model (GBL) for 6 weeks while the control group used a "problem based teaching" model (non-GBL). Pre-test/post-tests carried out. T-test and Levene's test used to analyse data.	Significant differences in learning motivation of the experimental group ( $F=7.655$ , $p=0.007<0.05$ ). No significant differences in learning achievement or reasoning ability between the groups ( $F=0.851$ , $p=0.358 > 0.05$ ). Overall high motivation affected children's learning and low motivation led to low productivity.

#### Appendix C: Perceptual and cognitive skill outcomes

Author(s)	Aims/Objectives of Study	Methods	Conclusions
Bottino, Ott, and Benigno (2009)	Identifying design and interface characteristics of digital mind games fostering reasoning and problem solving.	A three year longitudinal observation involving 40 participants (14M, 26F), aged 8-10 years (3 <sup>rd</sup> -5 <sup>th</sup> grade).	Characteristics fostering reasoning and problem solving abilities were adjustable difficulty, availability of hints and ability to sustain and orient cognitive activities. Cognitive overload was an obstacle to learning.
Cha, Baek and Xu (2008)	Exploring how gender, self-efficacy toward computers, logical thinking, and attitude toward gaming affect gaming achievement in GBL.	Participants were 72 5 <sup>th</sup> -grade students (39M and 33F, 12 years old). They had no previous experiences with the Zoombini game. Independent-sample t-test used to test gender differences in each variable and gaming achievement. Correlations among variables for all participants of each gender were calculated using Pearson's correlations coefficient.	GBL can be beneficial to players who are internally controlled and high in logical thinking when these variables are not part of the learning.
Ketamo (2009)	Applying teachable agents in educational games for children less than 12 years using Animal	<b>Experimental group 1</b> (N=59) for lab testing with elementary school teachers who like to use educational games as part of their teaching. Group	A clear causality between quality of the taught semantic networks in the game world and players' knowledge in real life and that learning away is an

	Class.	played Animal Class, a pre-school geometry game, or a 6 <sup>th</sup> -grade maths game. Group was tested empirically before and after playing the game with paper-based test. <b>Experimental group 2 (N=231)</b> to study the AI's capability to learn real-life situations with high ecological validity. Group 2 was observed only.	important feature when trying to enable conceptual change in educational games.
Linek and Marte (2011)	Evaluating the impact of background music in an educational adventure game on intrinsic motivation, the flow experience and cognitive load and learning.	A pre-test/post-test experimental/control group design. Utilised 59 school children (38M, 21F, M=13.6 years (SD=.89)). 27 in the experimental group (22M, 5F), 32 in the control (16M, 16F). 3 different schools in Paris. A two-group design was used to investigate the motivational and cognitive impact of background music with the presence (experimental) or absence (control) of background music. A t-test was utilised.	Background music had no effect on learning. No differences between the groups regarding grades, physics attitude, general music usage, and gaming-experience.
Miller and Robertson (2011)	Investigating effects of a COTS game on children's mental computation skills/self-perception.	Pre-test/post-test experimental/control group design utilised with 634 primary children (10–11 years old) from 32 Scottish schools. Schools were randomly assigned to experimental (used a games console for 20 minutes each day, running a 'brain training' game) or control conditions (normal routine). The treatment period was 9 weeks.	Significant gains in accuracy and speed of calculations were found in both groups. Gains in the experimental group were 50% greater in terms of accuracy, and twice those of the controls in speed. No significant changes in self-concept for either group. A statistically significant improvement in attitude towards school among the experimental group.
Robertson (2012)	Examining differences in characteristics of games produced by learners.	Field study in a class of 25 children aged 11-12 years, using Adventure Author.	Mean score of 49% for the games created. Females tended to score higher overall in all dimensions of the gaming assessment criteria. A higher variance of quality in the games produced by males. The game making process showed females spent more time undertaking story writing using peer review feedback to design games.

### Appendix D: Behavioural change outcomes

Author(s)	Aims/Objectives of Study	Methods	Conclusions
Dias and Agante (2011)	Investigating how healthy food stimuli can improve eating behaviour.	Study involved 231 participants (117M, 114F), grade 2 and 3 (7 to 8 years old), divided into 2 groups: group 1 played the healthy food advergame; group 2 played the less healthy food advergame. A $\chi^2$ test of association and t-tests were utilised.	Results show that children tend to choose according to what was being advertised in the game.
Sun, Wang and Chan (2011)	Investigated whether scaffolding tools can reduce player frustration in the game Professor Sudoku.	Pilot study with 12 5 <sup>th</sup> -grade students. Then recruited 213 students from 7 6 <sup>th</sup> -grade classes in an elementary school located in Hsinchu City, Taiwan. The final sample consisted of 18 in the frustration control group, 23 in the demonstration group, and 23 in the no-scaffold group. ANOVA, Chi-Squared, Correlations utilised.	ANOVA indicated significant between-group difference ( $F=4.520$ , $p<.05$ ). Evidence suggested that reliance on supportive tools diminishes learning opportunities.
Tsai (2008)	Investigating whether playing with/ caring for a virtual pet dog improves empathy and humane attitudes.	Pre-test/post-test experimental/control groups utilising a mixed-methods approach. 36 4 <sup>th</sup> and 5 <sup>th</sup> grade students utilised (19M, 17F). Paired t-test used to analyse the results.	Significant improvement in post-test in humane attitudes and empathy, with female students leading in both. Play and care positively correlated to humane attitude scores.
Vannini et al. (2009)	Investigating effects of a virtual learning strategy ("FearNot!") to increase coping skills and further heighten empathy in bullying victims.	Used a quasi-experimental, pre/post-test control group design. Twenty-six (year 5) primary school classes from the UK and 22 3 <sup>rd</sup> -grade primary school classes from Germany. Sample of 1,133 included 530 (47%) in the istudy and 603 (53%) in the control group; 642 (57%) pupils from the UK and 491 (43%) from Germany. The sample consisted of 587 (52%) boys and 546 (48%) girls, with an age range between 7 and 11 years ( $M=8.9$ ; $SD=0.7$ ; $n=9$ missing data). T-test, ANOVA, MANOVA used to analyse the results.	More UK than German children were involved in bullying as either bully and/or victims. More UK children and more girls were nominated by their classmates as defenders at baseline. Defenders were the most popular children within their classes.

- A large systematic literature review of computer games at Primary Education.
- Uses a multi-dimensional framework to ensure the quality of the studies analysed.
- Searches through 18,000 papers to identify 105 relevant papers in the field.
- Provides a snap-shot in time and repository in computer games for Primary Education.