Original article

doi: 10.1111/j.1365-2729.2012.00489.x

Journal of Computer Assisted Learning

Serious games as new educational tools: how effective are they? A meta-analysis of recent studies

C. Girard,* J. Ecalle* & A. Magnant†

*Laboratoire Etude des Mécanismes Cognitifs (EA 3082), Université Lyon (2), Bron Cédex, France
†Institut Universitaire de France, Paris, France

Abstract Computer-assisted learning is known to be an effective tool for improving learning in both adults and children. Recent years have seen the emergence of the so-called 'serious games (SGs)' that are flooding the educational games market. In this paper, the term 'serious games' is used to refer to video games (VGs) intended to serve a useful purpose. The objective was to review the results of experimental studies designed to examine the effectiveness of VGs and SGs on players' learning and engagement. After pointing out the varied nature of the obtained results and the impossibility of reaching any reliable conclusion concerning the effectiveness of VGs and SGs in learning, we stress the limitations of the existing literature and make a number of suggestions for future studies.

Keywords engagement, learning effect, serious game, video games.

Introduction

Background

In recent years, many new ways of teaching academic and professional skills to children and adults have been tested using multimedia technologies in the form of software products, educational computer games or video games (VGs) (Kebritchi *et al*. 2010; Lorant-Royer *et al*. 2010). Although the idea that games are effective learning tools is not new (Annetta *et al*. 2009), the question has only recently become a subject of experimental research. Some researchers claim that games permit constructive, situated and experiential learning, which is enhanced by active experimentation and immersion in the game (Squire 2008; Hainey *et al*. 2011). Their adopted perspective highlights the great advantage of games compared with traditional methods such as face-to-face or pencil-and-paper teaching. In addition, the traditional linear approach to learning seems to be counter-intuitive to many teachers and games could allow them to escape from the constraints it imposes (Tanes & Cemalcilar 2010).

The rapid growth of multimedia technologies over the last 20 years means that today's children and young adults were born in a computerized world and are used to handling all kinds of software products and games. They are known as the '*digital native*' or '*Net Generation*' (Prensky 2001; Westera *et al*. 2008; Annetta *et al*. 2009; Bekebrede *et al*. 2011). It therefore seems natural to think that this population will be very receptive to computer-based learning. Indeed, almost all the young people in the Net Generation greatly enjoy using computers and spend much of their time browsing or playing computer games. Digital learning will be enhanced by the large amount of time spent navigating in the software as well as by the high level of motivation and involvement in the activity displayed by learners (Boot *et al*. 2008; Tanes & Cemalcilar 2010; Wrzesien & Raya 2010). Moreover, it seems important to maintain a

Accepted: 25 March 2012
Correspondence: Jean Ecalle, Laboratoire Etude des Mécanismes Cognitifs (EA 3082), Université Lyon (2), 5 av, Mendès-France, 69676 Bron Cédex, France. Email: ecalle.jean@wanadoo.fr

level of continuity and consistency between the tools used in education and users' everyday home activities (Burnett 2009). Authors have listed many arguments in support of the effectiveness of computerized tools in learning: ease of accessibility, ease of updating content, low cost per person served, high level of interactivity, high level of user-specific customization, ability to use attractive graphics, and acceptability of the game by users as an engaging and entertaining activity (Beale *et al.* 2007; Boot *et al.* 2008; Westera *et al.* 2008; Burnett 2009; Kebritchi *et al.* 2010; Tanes & Cemalcilar 2010; Wrzesien & Raya 2010; Hainey *et al.* 2011).

Recently, new computerized tools known as 'serious games' (SGs) have appeared in the educational games market. By combining gaming and learning, SGs represent a new area of interest in the educational field. SGs are '*games primarily focused on education rather than entertainment*' (Miller *et al.* 2011, p. 1425). More specifically, SGs are VGs with a useful purpose, e.g. for training, education, knowledge acquisition, skills development, etc. We will discuss this point at greater length in section 'SGs vs. VGs' where we provide a precise definition of SGs and focus on how they differ from VGs. To prove the usefulness of SGs, researchers need to show that their specific characteristics (i.e. that from the design stage, they are intended to serve a useful purpose) make it possible to enhance learning in players. Although many authors seem convinced of the effectiveness of SGs (Johnson *et al.* 2005; Kiili 2005; de Freitas & Oliver 2006; Radillo 2009; Szilas & Sutter-Widmer 2009), the experimental evidence in favour of this position is meagre. In the educational field, as in medicine, the scientific assessment of training or learning effects usually takes the form of randomized controlled trials (Schulz *et al.* 2010). Indeed, this experimental design is considered to be the gold standard for the evaluation of both medical treatment and educational interventions. As Sibbald and Roland (1998) explain, '*randomized controlled trials are the most rigorous way of determining whether a cause-effect relation exists between treatment and outcome and for assessing the cost effectiveness of a treatment*' (p. 201). Our aim in the present paper was to review all experimental randomized controlled trial studies that have used VGs and/or SGs for learning in order to compare the effectiveness of these two types of digital application.

State of the literature on computer-assisted learning (CAL) and SGs

Reviews of the effectiveness of CAL and SGs

To place this review in context, we first summarized the existing literature on the effectiveness of CAL and reviewed the relevant papers devoted to SGs. A number of studies have attempted to prove that the digital devices used in teaching in recent years have a positive effect. First, many of these studies have examined the effectiveness of CAL and, overall, have concluded that these applications make it possible to enhance learning compared with traditional methods of teaching such as face-to-face lessons or pencil-and-paper studying. Blok *et al.* (2002) analysed 42 published quantitative studies in English and Dutch on the effectiveness of using computers to teach beginning reading to children aged 5–12. Blok *et al.*'s meta-analysis found a small positive effect of computer-assisted beginning reading instruction compared with traditional instruction. Other authors have highlighted the effectiveness of CAL compared with conventional teaching methods both in the same (see Mioduser *et al.* 2000, for a study of the learning value of computer-based instruction of early reading skills) and other fields (see Schittek *et al.* 2001, for a review of CAL vs. traditional teaching methods in the field of dentistry).

In a meta-analysis of the effect of computer games and interactive simulations on learning, Vogel *et al.* (2006) reported that these two digital applications lead to better cognitive outcomes and that the students prefer them to traditional teaching methods. However, it is not clear from the article exactly what types of application were considered (CAL, VG, SG or others). Additionally, this study has been criticized for having reviewed a large number of unpublished studies (Sitzmann 2011). More recently, Mikropoulos and Natsis (2010) examined educational virtual environments (EVEs) and produced a review of empirical research into these applications undertaken over the previous 10 years. They showed that EVEs are effective teaching tools because they promote constructivist learning.

However, to date there are no summary reviews of SGs and their effectiveness. A number of researchers have published reviews of specific SGs, such as Baranowski *et al.* (2008) in the field of health-related behaviour change and Goh *et al.* (2008) who studied psychotherapeutic gaming interventions. The latter examined the game-based training of children and adolescents with autistic spectrum disorders, anxiety disorders or language learning impairments. They concluded that training can be effective in the field of mental health treatment and discussed the game design characteristics and human factors (age, gender, socio-economic status, and level of cognitive and emotional development) that can affect the educational effectiveness of games. They drew up a table summarizing the types of game that are the most suitable in the light of the players' mental health and the objectives of the treatment. For example, they recommended using simulation games and virtual reality software to induce behaviour changes in subjects with phobic disorders. Baranowski *et al*. (2008) examined 27 studies of VGs designed to induce health-related behaviour change. Given the criteria used for exclusion from and inclusion in their review, it seems to us that most of the games they considered were SGs (and not VGs as the authors claim). Their analysis showed that the studies vary greatly on the dimensions of game design, game purpose, game features and measures. Their overall finding is that the use of SGs to promote health-related behaviour change has a positive effect (except in one study). The researchers identified important criteria that enhance the effectiveness of SGs (story genre, immersion, fantasy, design and gameplay) and stressed the need to perform further research in the field of SGs targeting health-related behaviour change. The main limitation of Goh *et al*.'s study lies in its lack of precision concerning the way the examined games were selected, which means that we do not know whether they consisted of VGs, SGs, EVEs or CAL. The presentation of the inclusion and exclusion criteria is not sufficiently clear and no accurate description of the games is given. Moreover, the experimental studies examined in the review seem to be too old (from 1998 to 2005) to include SGs. The second limitation of this review is that only games relating to a specific field (psychotherapeutic gaming interventions) were considered. This also constitutes a significant limitation to Baranowski *et al*.'s study, which compounds the considerable variability of the examined studies and the confusion between VGs and SGs. It is therefore difficult to generalize on the basis of the positive outcomes of the serious gaming interventions examined in these two reviews.

The main limitation affecting attempts to review studies of the effectiveness of a given type of educational tool (ludo-educative applications, SGs, simulations, etc.) lies in the lack of precision in the definition of the category of tool studied. We therefore believe that the first step in any review of this type should consist of a clear and rigorous definition of the characteristics, which a tool must possess in order to be considered to belong to a specific studied category. For example, Sitzmann (2011) published a well-documented meta-analysis of 55 research reports relating to the instructional effectiveness of simulation games. The author clearly defined the term 'simulation game' to indicate what types of game were included in or excluded from the analysis. She then performed a review of the literature in order to determine the effect of simulation games, relative to a comparison group, on three affective (motivation, trainee reactions and self-efficacy), one behavioural (effort), two cognitive (declarative knowledge and retention) and two skill-based (procedural knowledge and transfer) training outcomes. She concluded that technology can enhance learning but added that '*technology is a means for delivering teaching but does not have a direct effect on learning*' (Sitzmann 2011, p. 516). Furthermore, she outlined the importance of certain positive factors for improving learning during training using simulation games: integration of game use within a programme of instruction, high level of activity on the part of the learner, importance of the debriefing session after game play and no limitation to the time available to play the game. At first sight, this study seems very well documented and the author's approach is very rigorous at both the theoretical and statistical levels (calculation of effect sizes, confidence intervals and other statistical analyses). However, a closer inspection reveals that the 'simulation games' selected for this meta-analysis were not equivalent and do not fit with our definition of 'simulation games' or 'SGs'. The author included in her analysis games which are too old to be simulation games, games which do not resemble modern simulation games, games which do not meet the criteria necessary in order (according to us) to be categorized as simulation games and games which have no ludic content whatsoever. The majority of the games reviewed by Sitzmann (2011) should be considered as educational applications or CAL but not simulation games. This is why it is important to clearly define and present criteria for the selection of the reviewed studies.

Other publications relating to SGs: definition, design and characteristics

Because our aim here is to review all relevant scientific publications relating to VGs and SGs, we first summarize a varied selection of important non-experimental works before presenting our review of the experimental studies. Raybourn (2007) proposed a method for designing SG-based adaptive training systems and stressed the lack of empirical studies showing the effect of training with SGs on learning and transfer, as, too, did Annetta *et al*. (2009), Wrzesien and Raya (2010) and Hainey *et al*. (2011). To provide an overview of the development of educational tools, Martin *et al*. (2011) analysed the evolution of technology trends from 2004 to 2014 (data and predictions) and considered (among other things) immersive environments such as games and virtual worlds. These data are purely bibliometric in nature and provide no information about the contents of the applications. The result of interest for our purposes here is that immersive environments (which include SGs) have been developed since 2007. Adopting a highly theoretical approach, Marsh (2011) defined and traced the history of the expression 'serious game', reassessed the name (Is it really a game? Is it really serious?) and finally stated the characteristics of SGs, while also giving many examples of existing SGs. Marsh's publication is not a review of SGs, Instead, the author wanted to propose a categorization and definition of SGs, which could be agreed on by all researchers, developers, academics, etc. working in the area of SGs. This should simplify the design and evaluation of SGs. Many researchers have tried to define the characteristics of SGs (Johnson *et al*. 2005; Raybourn 2007; Goh *et al*. 2008; Muratet *et al*. 2009; Marsh 2011) and placed the emphasis on the importance of gameplay, feedback, human–computer interaction, challenge, scenario, fun, immersion, game design and learning–game integration. Other authors have attempted to define frameworks or models for the design of SGs (Kiili 2005; de Freitas & Oliver 2006; Westera *et al*. 2008; Szilas & Sutter-Widmer 2009), described the SG production process (Marfisi-Schottman *et al*. 2009) or identified various problems and questions relating to SGs (Rebetez & Betrancourt 2007). However, very few empirical studies of SGs have as yet been performed and no summary of these works exists. That is why it is now so important to review the experimental studies of VG and SG-based training conducted over the last 5 years (in particular in the light of the new technology trends in education as described by Martin *et al*. 2011). Given our concerns regarding categorization, we first define the features of what we consider to constitute a VG, on the one hand, and a SG, on the other. We then define certain necessary inclusion criteria concerning the methodology of the experimental studies (randomized controlled trials, type of measures). Finally, we analyse the results of the selected studies and attempt to interpret them in the light of certain reliable theories of engagement and motivation, as well as consider the limitations and the outcomes of these studies.

SGs vs. VGs

After considering all the publications cited above, we decided to adopt a part of the definition that Marsh (2011) gives for SGs: '*Serious games are digital games, simulations, virtual environments and mixed reality/media that provide opportunities to engage in activities through responsive narrative/story, gameplay or encounters to inform, influence, for well-being, and/or experience to convey meaning*.' (Marsh 2011, p. 63). For us, the only difference between a SG and a VG lies in their intended purpose: usefulness for the former, entertainment for the latter. That is to say, SGs are VGs with a useful purpose. This idea is not shared by Marsh (2011), who argued that a VG used for a useful objective can also be considered to be a SG. We do not accept this viewpoint as such an argument could lead us to consider every VG to be a SG because it can always be claimed that the objective of training with a VG is to improve visuospatial or language skills (if playing the VG requires the reading of certain texts). We therefore maintain that the utility of purpose has to be present from the outset (right from the very first step in the design of the SG) and not added to the game subsequently.

It is, moreover, important to mention the existence of a huge variety of SGs (as illustrated in Supporting Information Table S1) that differ both in terms of the type of game (strategy game, brain training, adventure game, etc.) and the type of skill trained (health, military, academic knowledge, etc.). It should also be noted that some VGs or SGs can be categorized into more than one type. This enormous variety of VGs and SGs does not simplify the task of analysing their effectiveness in learning.

Methodology

Formulation of the problem

As we emphasize in our 'Introduction' section, although many researchers are convinced of the effectiveness of SGs in learning, there is a lack of clear and pertinent empirical evidence in support of this viewpoint (Raybourn 2007; Farrington 2011). For the purposes of this review, we attempted to identify all the experimental studies that have used SGs for training or learning and assessed their results in terms of both effectiveness and acceptability. To highlight the advantage/benefit of SGs over VGs and thus justify according them a specific status, we also considered experimental studies that have used VGs for training or learning. We compared the results obtained with SGs with those found with VGs.

Data collection and selection

We performed searches in referenced databases such as ScienceDirect and PubMed for articles published from 2007 to 2011 in scientific journals or as proceedings of conferences and symposia and relating primarily to the fields of cognitive science, psychology, human–computer interaction and education, but also other scientific fields such as medicine or engineering in which training has been performed using SGs or VGs.

As keywords, we used strings consisting of ('digital game' or 'computer game' or 'serious game' or 'video games' or 'game-based learning') and ('learning' or 'training' or 'teaching' or 'effects' or 'education'), and also searched for other articles of interest cited in the papers that we selected.

We first of all selected 30 publications from 15 different journals (see Table 1). We then performed a second selection excluding all the studies that did not use SGs or VGs (Duque *et al*. 2008; Ke 2008; Lorant-Royer *et al*. 2008; Sung *et al*. 2008; Papastergiou 2009; Robertson & Miller 2009; Baker *et al*. 2010; Bloomfield *et al*. 2010; Liu & Chu 2010; Lindström *et al*. 2011; Brom *et al*. 2011). To do this, we defined necessary criteria that are shared by SGs and VGs as well as specific and necessary criteria determining categorization as either a SG or a VG, on the basis of various theoretical articles (Johnson *et al*. 2005; Kiili 2005; de Freitas & Oliver 2006; Baranowski *et al*. 2008; Marfisi-Schottman *et al*. 2009; Radillo 2009; Szilas & Sutter-Widmer 2009; Marsh 2011). Supporting Information Table S1 presents the shared and specific criteria of SGs and VGs together with justifications for their application. For each game in the first selection, we specify the presence or absence of each necessary or specific criterion justifying categorization as a VG or SG and indicate why the corresponding criteria were chosen.

Because we wanted to keep only randomized controlled trial studies, we defined the inclusion and exclusion criteria to permit us to identify the relevant experimental studies. As a result, the following are excluded from the present review: articles that do not present empirical data derived from the evaluation of SGs or VGs (Radillo 2009; Kobes *et al*. 2010), studies that do not have at least a 'pre-test – game – post-test' design (Ciavarro *et al*. 2008; Coller & Scott 2009; Kim *et al*. 2009), studies that do not use quantitative measures (Lim 2008; Watson *et al*. 2011) or in which no control group is included (Tuzun *et al*. 2009; Echeverrìa *et al*. 2011; Miller *et al*. 2011).

Application of these inclusion and exclusion criteria resulted in a total of nine studies published between 2007 and 2011 in six different journals being included in this review. Of these studies, two tested SGs or VGs in the field of cognition (Boot *et al*. 2008; Lorant-Royer *et al*. 2010), four tested SGs or VGs used in the academic field (Annetta *et al*. 2009; Kebritchi *et al*. 2010; Wrzesien & Raya 2010; Hainey *et al*. 2011), two in the medical field (Beale *et al*. 2007; Knight *et al*. 2010) and one in the social field (Tanes & Cemalcilar 2010). Because some of the studies tested several different games, the total number of studied games covered by this review is 11, namely six SGs and five VGs.

Data analysis

We examined the selected studies from various perspectives. First, we analysed the characteristics of the tested populations: number of subjects in the experimental group and in the control group, minimum and maximum ages of the subjects, and specific characteristics of the subjects. We then analysed the experimental methodology used in the studies: type of measures used in the pre-test and the post-test, duration of training with the game,

Table 1. Result of the selection of the first 30 studies: authors, year and journal of publication.

Authors	Year	Journal
[1]Annetta, Minogue, Holmes & Cheng	2009	Computers & Education
[1]Beale, Kato, Marin-Bowling, Guthrie & Cole	2007	Journal of Adolescent Health
[1]Boot, Kramer, Simons, Fabiani & Gratton	2008	Acta Psychologica
[1]Hainey, Connolly, Stansfield & Boyle	2011	Computers & Education
[1]Kebritchi, Hirumi & Bai	2010	Computers & Education
[1]Knight, Carley, Tregunna, Jarvis, Smithies, de Freitas, Dunwell & Mackway-Jones	2010	Resuscitation
[1]Lorant-Royer, Munch, Mesclé & Lieury	2010	European Review of Applied Psychology
[1]Tanes & Cemalcilar	2010	Journal of Adolescence
[1]Wrzesien & Raya	2010	Computers & Education
Baker, D'Mello, Rodrigo & Graesser	2010	International Journal of Human-Computer Studies
Bloomfield, Roberts & While	2010	International Journal of Nursing Studies
Brom, Preuss & Klement	2011	Computers & Education
Ciavarro, Dobson & Goodman	2008	Computers in Human Behavior
Coller & Scott	2009	Computers & Education
Duque, Fung, Mallet, Posel & Fleiszer	2008	Journal of the American Geriatrics Society
Echeverrìa, Garcìa-Campo, Nussbaum, Gil, Villalta, Améstica & Echeverrìa	2011	Computers & Education
Ke	2008	Computers & Education
Kim, Park & Baek	2009	Computers & Education
Kobes, Helsloot, de Vries & Post	2010	Procedia Engineering
Lim	2008	Computers & Education
Lindström, Gulz, Haake & Sjödén	2011	Journal of Computer Assisted Learning
Liu & Chu	2010	Computers & Education
Lorant-Royer, Spiess, Goncalves & Lieury	2008	Bulletin de psychologie
Miller, Chang, Wang, Beier & Klisch	2011	Computers & Education
Papastergiou	2009	Computers & Education
Radillo	2009	Enfances & Psy
Robertson & Miller	2009	Procedia social and behavioral sciences
Sung, Chang & Huang	2008	Computers in Human Behavior
Tuzun, Yilmaz-Soylu, Karakus, Inal & Kizilkaya	2009	Computers & Education
Watson, Mong & Harris	2011	Computers & Education

[1]Final nine selected studies.

and type of VG or SG. Finally, we analysed the results of the experimental studies in the light of the following key question: did training with the game (or games) have a positive effect, a negative effect or no effect?

Results

Population characteristics

The populations tested in the selected studies were large enough to possess a high level of scientific statistical validity, ranging from 48 (Wrzesien & Raya 2010) to 371 subjects (Beale *et al*. 2007), with an average of 149 (SD = 104) subjects for each of the nine selected studies.

On average, there were 80 (SD = 51) subjects in the experimental groups and 69 (SD = 59) in the control groups (based on the authors' designations of the groups). It is important to emphasize the ambiguity of the term 'control group'. Indeed, some authors considered subjects who received no training to constitute the control group (Boot *et al*. 2008; Lorant-Royer *et al*. 2010; Tanes & Cemalcilar 2010), whereas others trained the subjects in their control groups using pencil-and-paper backup material (Annetta *et al*. 2010; Knight *et al*. 2010; Hainey *et al*. 2011) or traditional learning methods (Kebritchi *et al*. 2010; Wrzesien & Raya 2010) and yet others presented their control groups with a non-tested VG (Beale *et al*. 2007). We shall return to this point later (see section 'The "control group" problem').

As far as the age of the studied populations is concerned, the subjects were aged between 9 (Lorant-Royer *et al*. 2010) and 47 years (Hainey *et al*. 2011), and most of the studies focused on adolescents and young students. Almost all of the studies used a mixed population of subjects within a well-balanced design, with the exception of Boot *et al*. (2008), who tested 75 females and seven males.

In some of the studies, the population was selected on the basis of specific characteristics. The subjects in Boot *et al*.'s (2008) study were non-gamers (playing less than 1 h of VGs a week for the last 2 years), those in Tanes and Cemalcilar (2010) had never played SimCity or The Sims, and the subjects in Knight *et al*. (2010) were students attending Major Incident Medical Management and Support Courses. Finally, the subjects in Beale's (2007) study were patients diagnosed as having cancer, who had been receiving treatment for at least 4–6 months and were still being treated at the time of the study.

Experimental methodology and procedure

Almost all the selected studies used a classical 'pre-test/training/post-test' experimental design, with the exception of two studies in which there was no pre-test (Annetta *et al*. 2009; Knight *et al*. 2010). Although these latter studies therefore failed to meet one of the criteria for inclusion in the present review, we retained these two papers because the groups were clearly balanced in terms of their initial level of knowledge and because these studies seemed interesting from other points of view (see below). It should also be noted that some of the selected studies (Beale *et al*. 2007; Boot *et al*. 2008) added a second post-test in their experiments.

Because the studies addressed different fields, the assessed skills were consequently equally diverse: academic knowledge (Annetta *et al*. 2010; Kebritchi *et al*. 2010; Wrzesien & Raya 2010), cognitive skills (Boot *et al*. 2008; Lorant-Royer *et al*. 2010), technical professional knowledge (Knight *et al*. 2010; Hainey *et al*. 2011) or other specific knowledge such as knowledge about cancer therapies (Beale *et al*. 2007) and the perception of urban issues (Tanes & Cemalcilar 2010).

Three kinds of measure were used in these studies: knowledge questionnaires, academic tests and cognitive tests. In some of the studies, the authors also investigated the subjects' level of engagement, motivation or satisfaction by means of post-test questionnaires (Annetta *et al*. 2010; Kebritchi *et al*. 2010; Wrzesien & Raya 2010; Hainey *et al*. 2011).

The time spent playing the game differed according to the studies: some of the studied populations interacted with game during a single session lasting between 15 and 90 min, whereas others played for several weeks or months (one or two sessions lasting 1 or 2 h each week).

Measures of learning effect and engagement

In the selected studies, the effect of training on learning (acquisition of skills or knowledge) was measured by calculating the difference between the pre-test and the post-test scores on the questionnaires or cognitive tests. The authors compared the increases in scores observed in the control and experimental groups in order to assess the effectiveness of using the tested games (VG or SG or both).

To sum up the results, three of the 11 studied games had a positive effect on learning compared with other types of training or no training at all. These included two SGs (Re-Mission and DimensionM) and one VG (SimCity). Seven games had no beneficial effect on learning, including three SGs (MEGA, RCAG, E-Junior) and four VGs (Indiana Jones & the Emperor's Tomb, Medal of Honor Allied Assault, Rise of Nations, New Super Mario Bros.). Finally, the results for one SG (Triage Trainer) were mixed (see Table 2). Because of the lack of precise quantitative data in many studies, it was not possible to calculate the size of the learning effect.

The data concerning the subjects' ratings revealed that two SGs (MEGA and E-Junior) aroused a higher level of engagement and motivation than traditional teaching methods, but that the SG DimensionM had no beneficial effect on motivation.

Discussion

Effectiveness of training with VGs and SGs for learning

The varying nature of the results obtained in the small number of experimental studies selected on the basis of the inclusion and exclusion criteria used in the present

Table 2. Results of the final nine selected studies for each serious game (SG) and video game (VG): learning effect and effect of the game on engagement.

Authors	SG or VG	Name of the game	Learning effect of the game	Engagement effect of the game
Annetta, Minogue, Holmes & Cheng	SG	(no name) MEGA	No	Yes
Beale, Kato, Marin-Bowling, Guthrie & Cole	SG	Re-Mission	Yes	?
Beale, Kato, Marin-Bowling, Guthrie & Cole	VG	Indiana Jones & The Emperor's Tomb	No	?
Boot, Kramer, Simons, Fabiani & Gratton	VG	Medal of Honor Allied Assault	No	?
Boot, Kramer, Simons, Fabiani & Gratton	VG	Rise of Nations	No	?
Hainey, Connolly, Stansfield & Boyle	SG	RCAG	No	?
Kebritchi, Hirumi & Bai	SG	DimensionM	Yes	No
Knight, Carley, Tregunna, Jarvis, Smithies, De Freitas, Dunwell & Mackway-Jones	SG	Triage Trainer	Yes/no	?
Lorant-Royer, Munch, Mesclé & Lieury	VG	New Super Mario Bros™	No	?
Tanes & Cemalcilar	VG	SimCity	Yes	?
Wrzesien & Raya	Serious virtual world	E-junior	No	Yes

review means that the effectiveness of VGs and SGs remains to be proven. Indeed, only a few of the games resulted in improved learning, with the others having no positive effect on knowledge and skills acquisition when compared with more traditional methods of teaching, on the one hand, or to a control group which received no training, on the other.

Although it is not possible to conclude that SGs are more effective to VGs on the basis of only nine studies, it is nevertheless both possible and very interesting to use the results of this meta-analysis to provide a number of hypothetical explanations concerning the limitations of the various existing studies (see section 'Limitations of the selected empirical studies').

The capability of VGs and SGs to engage and motivate subjects

The results relating to the engagement and motivation of players of SGs and VGs reported in the selected studies cannot be generalized because of the limited data available. However, previous studies seem to confirm that subjects find games more motivating and engaging than traditional methods such as pencil-and-paper study or face-to-face teaching.

According to Prensky's (2001) Net Generation concept, subjects are more willing to spend time training with VGs and SGs because they are used to and enjoy playing, given that VGs have been part of their everyday lives since a young age. In other words, people today are fascinated and stimulated by VGs (in the broad sense), which are engaging and entertaining to play (Johnson *et al*. 2005; Westera *et al*. 2008; Wrzesien & Raya 2010). They are therefore more motivated by game-based than by traditional learning methods (Papastergiou 2009; Radillo 2009; Annetta *et al*. 2010).

Limitations of the selected empirical studies

As mentioned above, it is difficult to draw reliable conclusions about the effectiveness of VGs and SGs because of the very varied nature of the studies (fields of research, experimental procedures, etc.) and because of their limitations, as we shall see below.

The 'control group' problem

The main limitation not only of the empirical studies selected for this review but also of the majority of other studies of SGs and VGs lies in the inconsistent use of the concept of 'control group'. Depending on the authors, the control group may be a group of subjects who receive no training, subjects who are trained using traditional methods (pencil-and-paper or face-to-face teaching) or a group trained using a different game in order to study learning effectiveness compared with another VG or SG. It is therefore impossible to compare the results of the effectiveness of learning in the different studies because there is no common baseline for such a comparison.

This raises the question: is there an ideal control group? On the one hand, it seems important to compare the learning effect in the group playing the tested game with that observed in a group that receives no training in order to show that the benefit of the training activity is not due simply to automatic implicit learning processes, a test–retest effect or a learning effect because of academic teaching (Lorant-Royer *et al*. 2010). On the other hand, it might seem fairer (for subjects who are students in the same class, for example) and more valid (methodologically) to compare two (or more) types of training (for example, pencil-and-paper vs. game) in order to assess which is the most effective for learning (Kebritchi *et al*. 2010). Indeed, it might seem logical that any type of training would be more effective than no training (even if this is not always the case) and that the best way to prove the effectiveness of a given type of teaching tool is to compare it with another teaching tool with the same educational content.

In our opinion, the best way to assess and prove the effectiveness of any given type of training (VG, SG or other) is to compare it with at least a group that receives no training and a group trained using a different type of training material.

Transfer of acquired knowledge and skills

Any discussion of the effectiveness of training using VGs and SGs must also address the question of the transfer of acquired knowledge and skills. Very few of the selected studies assessed the effect of game-based training on the subjects' everyday lives. It would seem to be very important to find out, in particular in the studies in which a training effect was observed, whether the knowledge acquired during the game-based training persists in the long term and is useful in real-life situations. For example, in the study conducted by Beale *et al*. (2007), it would be interesting to know whether the patients trained using Re-Mission (and who exhibited improved knowledge after training) changed their behaviour regarding their treatment and everyday health concerns. Similarly, it would be interesting to assess the consequences of game-based training on academic outcomes at the end of the school year in the studies conducted by Annetta *et al*. (2010), Kebritchi *et al*. (2010) and Wrzesien and Raya (2010).

The question of the transfer of acquired knowledge and skills is also addressed by Boot *et al*. (2008), who observed no transfer of cognitive skills in their experimental study. According to Lorant-Royer *et al*. (2008), Serruya and Kahana (2008) and Lorant-Royer *et al*. (2010), training is usually not sufficiently specific and, consequently, has little impact on real life (Farrington 2011). This is a critical point that has to be addressed by research into the effectiveness of game-based training.

The various different types of SGs

As already mentioned in section 'SGs vs. VGs' in which we introduced the definition of 'SG', it is very difficult to speak about 'SGs' in general terms because there is a huge variety of SGs, both in terms of type of game and the skills that are trained. In consequence, it is impossible to draw any general conclusion about the effectiveness of SGs and each type of SG has to be considered specifically and at an individual level. That is why, in any experimental study of the effectiveness of a SG, it is necessary to clarify all the key parameters of the game, as the quality of the design, all elements of the game's mechanics, the devices used during play, the game scenario, the integration of useful elements in the gameplay, the context of use, etc.

What evidence suggests that SGs are effective for learning?

Given that, as we have explained, the effectiveness of SGs has not as yet been proven and that the existing studies suffer from serious limitations, it is natural to ask why many researchers are convinced of the power of SGs for learning (Farrington 2011). In this section, we try to explain why, despite the lack of experimental evidence, there are reasons to believe in the effectiveness of SGs for learning.

Researchers often claim that there is a causal relationship between engagement, subjects' motivation and their learning. Annetta *et al*. (2010) suggested that cognitive engagement in the training, coupled with affective engagement and motivation (Baker *et al*.

2010; Knight *et al.* 2010; Sitzmann 2011), can be a factor that impacts the effectiveness of learning. Furthermore, the fact that subjects are engaged and motivated by the game makes them train for longer than with traditional materials and therefore makes a positive contribution to their progress (Annetta *et al.* 2009). One experimental study of educational games (Papastergiou 2009) compared a non-game version and a game version of two equivalent applications (content). The results showed that the students in the experimental group not only considered the game-based application to be more attractive and more educationally effective, but also acquired more knowledge than the students in the control group. The authors concluded that the game version of the application was more motivating and consequently more effective for learning. According to Wrzesien and Raya (2010), games like VGs and SGs boost intrinsic motivation in players (desire for challenging, independent mastery and curiosity). Players are consequently more engaged in the learning process and learn more than they do using traditional teaching methods. In their experimental study, these authors found no difference in learning between the two groups (experimental vs. control) despite the fact that the feedback questionnaires and observations revealed that the children in the serious virtual world environment were more motivated, more satisfied and more engaged than those in the traditional learning group. Wrzesien and Raya (2010) assumed that the innovative and entertaining features of the game distracted the children from their learning, but postulated that learning would become possible again once the novelty of the method had worn off. A number of authors concur in thinking that 3D elements and sophisticated game features might impair the subjects' engagement in the learning activity. This observation emphasizes why it is vital to achieve a compromise between stimulating ways of engaging the players through the features made available by the game and keeping their attention focused on learning (Annetta *et al.* 2010). Overall, authors agree on the beneficial effect on learning of the engagement and motivation generated by games. They often refer to the Flow Theory of Csikszentmihalyi (see Kiili 2005 for a detailed presentation). According to this theory, the 'flow' state experienced by players (state of complete involvement or engagement in an activity that results in an optimum experience of the activity) during the game has a positive effect on their learning (Sitzmann 2011). Good games are particularly adept at keeping subjects in a state of flow by increasing the skill level involved in the game as the skill level exhibited by the player increases (Kiili 2005). However, our observations indicate that not all the experimental results support this theory and that further research is therefore required.

Conclusion

Overall, the results of the nine studies analysed in this review and the arguments taken from the literature concerning the effect of engagement lead us to think that SGs might be powerful tools for learning. However, as emphasized by Annetta *et al.* (2009), Wrzesien and Raya (2010) and Hainey *et al.* (2011), and as we have seen in this review, there is a clear lack of empirical studies investigating the effectiveness of SGs in learning. In our opinion, it will be necessary to conduct many more experimental studies comparing the effect on learning of no training, pencil-and-paper training, VG training and SG training, before it is possible to claim that SG-based training is effective (Farrington 2011). It will also be necessary to perform longitudinal studies to assess the long-term effect of training using SGs. As Lorant-Royer *et al.* (2010) have suggested, we should avoid becoming overenthusiastic about the SGs that are currently flooding the market until their effectiveness for learning has been scientifically demonstrated. However, this review also needs to stress the beneficial aspects of the development of SGs, which could prove to be the educational tools of the future.

As stated above, today's young adults, adolescents and children are very used to and motivated by VGs and educational applications (Lorant-Royer *et al.* 2010). Consequently, playing this type of game could constitute a support for learning when traditional teaching methods are too boring (Wrzesien & Raya 2010). Annetta *et al.* (2009) suggest that they could provide a way of enhancing engagement and motivation in children with learning difficulties or attention disorders, for example. Moreover, the use of SGs permits the harmless simulation (Westera *et al.* 2008; Farrington 2011) of many physical situations and natural phenomena, which cannot be produced in real-world situations (such as ecological disasters, the spread of diseases, astronomical phenomena and emergencies; see, for example, Kobes *et al.* 2010). It also allows learners to explore environments that are inaccessible to most people (such as the ocean floor or outer space) and gives concrete form to certain abstract problems such as mathematical equations, thus opening up new learning possibilities for students. These potentials illustrate why it is important to continue to study the effectiveness of SGs on engagement and the associated learning effects.

References

Annetta L.A., Minogue J., Holmes S.Y. & Cheng M.-T. (2009) Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education* **53**, 74–85.

Baker R.S.J.D., D'Mello S.K., Rodrigo M.M.T. & Graesser A.C. (2010) Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies* **68**, 223–241.

Baranowski T., Buday R., Thompson D.I. & Baranowski J. (2008) Video games and stories for health-related behavior change. *American Journal of Preventive Medicine* **34**, 74–82.

Beale I.L., Kato P.M., Marin-Bowling V.M., Guthrie N. & Cole S.W. (2007) Improvement in cancer-related knowledge following use of a psychoeducational video game for adolescents and young adults with cancer. *The Journal of Adolescent Health* **41**, 263–270.

Bekebrede G., Warmelink H.J.G. & Mayer I.S. (2011) Reviewing the need for gaming in education to accommodate the net generation. *Computers & Education* **57**, 1521–1529.

Blok H., Oostdam R., Otter M. & Overmaat M. (2002) Computer-assisted instruction in support of beginning reading instruction: a review. *Review of Educational Research* **72**, 101–130.

Bloomfield J., Roberts J. & While A. (2010) The effect of computer-assisted learning versus conventional teaching methods on the acquisition and retention of handwashing theory and skills in pre-qualification nursing students: a randomised controlled trial. *International Journal of Nursing Studies* **47**, 287–294.

Boot W.R., Kramer A.F., Simons D.J., Fabiani M. & Gratton G. (2008) The effects of video game playing on attention, memory, and executive control. *Acta Psychologica* **129**, 387–398.

Brom C., Preuss M. & Klement D. (2011) Are educational computer micro-games engaging and effective for knowledge acquisition at high-schools? A quasi-experimental study. *Computers & Education* **57**, 1971–1988.

Burnett C. (2009) Research into literacy and technology in primary classrooms: an exploration of understandings generated by recent studies. *Journal of Research in Reading* **32**, 22–37.

Ciavarro C., Dobson M. & Goodman D. (2008) Implicit learning as a design strategy for learning games: Alert Hockey. *Computers in Human Behavior* **24**, 2862–2872.

Coller B.D. & Scott M.J. (2009) Effectiveness of using a video game to teach a course in mechanical engineering. *Computers & Education* **53**, 900–912.

Duque G., Fung S., Mallet L., Posel N. & Fleiszer D. (2008) Learning while having fun: the use of video gaming to teach geriatric house calls to medical students. *Journal of the American Geriatrics Society* **56**, 1328–1332.

Echeverrìa A., Garcìa-Campo C., Nussbaum M., Gil F., Villalta M., Améstica M. & Echeverrìa S. (2011) A framework for the design and integration of collaborative classroom games. *Computers & Education* **57**, 1127–1136.

Farrington J. (2011) From the research: myths worth dispelling: seriously, the game is up. *Performance Improvement Quarterly* **24**, 105–110.

de Freitas S. & Oliver M. (2006) How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education* **46**, 249–264.

Goh D.H., Ang R.P. & Chern Tan H. (2008) Strategies for designing effective psychotherapeutic gaming interventions for children and adolescents. *Computers in Human Behavior* **24**, 2217–2235.

Hainey T., Connolly T., Stansfield M. & Boyle E.A. (2011) Evaluation of a game to teach requirements collection and analysis in software engineering at tertiary education level. *Computers & Education* **56**, 21–35.

Johnson W.L., Vilhjalmsson H. & Marsella S. (2005) Serious games for language learning: how much game, how much AI? *Proceeding of The 2005 Conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology*, pp. 306–313. IOS Press, Amsterdam, The Netherlands.

Ke F. (2008) A case study of computer gaming for math: engaged learning from gameplay? *Computers & Education* **51**, 1609–1620.

Kebritchi M., Hirumi A. & Bai H. (2010) The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education* **55**, 427–443.

Kiili K. (2005) Digital game-based learning: towards an experiential gaming model. *The Internet and Higher Education* **8**, 13–24.

Kim B., Park H. & Baek Y. (2009) Not just fun, but serious strategies: using meta-cognitive strategies in game-based learning. *Computers & Education* **52**, 800–810.

Knight J.F., Carley S., Tregunna B., Jarvis S., Smithies R., de Freitas S., Dunwell I. & Mackway-Jones K. (2010) Serious gaming technology in major incident triage training: a pragmatic controlled trial. *Resuscitation* **81**, 1175–1179.

Kobes M., Helsloot I., de Vries B. & Post J. (2010) Exit choice, (pre-)movement time and (pre-)evacuation behaviour in hotel fire evacuation – behavioural analysis and validation of the use of serious gaming in experimental research. *Procedia Engineering* **3**, 37–51.

Lim C.P. (2008) Global citizenship education, school curriculum and games: learning Mathematics, English and Science as a global citizen. *Computers & Education* **51**, 1073–1093.

Lindström P., Gulz A., Haake M. & Sjödén B. (2011) Matching and mismatching between the pedagogical design principles of a math game and the actual practices of play. *Journal of Computer Assisted Learning* **27**, 90–102.

Liu T.-Y. & Chu Y.-L. (2010) Using ubiquitous games in an English listening and speaking course: impact on learning outcomes and motivation. *Computers & Education* **55**, 630–643.

Lorant-Royer S., Munch C., Mesclé H. & Lieury A. (2010) Kawashima vs. 'Super Mario'! Should a game be serious in order to stimulate cognitive aptitudes? *European Review of Applied Psychology* **60**, 221–232.

Lorant-Royer S., Spiess V., Goncalves J. & Lieury A. (2008) Programmes d'entraînement cérébral et performances cognitives: efficacité, motivation . . . ou «marketing»? De la Gym-Cerveau au programme du Dr Kawashima . . . (Brain training programs and cognitive performance: efficacity, motivation . . . or «marketing»? From Gym-Cerveau to Kawashima's brain training . . .). *Bulletin de Psychologie* **61**, 531–549.

Marfisi-Schottman I., Sghaier A., George S., Tarpin-Bernard F. & Prévôt P. (2009) Towards industrialized conception and production of serious games. *Proceeding of The International Conference on Technology and Education*, pp. 1016–1020. Paris, France.

Marsh T. (2011) Serious games continuum: between games for purpose and experiential environments for purpose. *Entertainment Computing* **2**, 61–68.

Martin S., Diaz G., Sancristobal E., Gil R., Castro M. & Peire J. (2011) New technology trends in education: seven years of forecasts and convergence. *Computers & Education* **57**, 1893–1906.

Mikropoulos T. & Natsis A. (2010) Educational virtual environments: a ten-year review of empirical research (1999–2009). *Computers & Education* **56**, 769–780.

Miller L.M., Chang C.-I., Wang S., Beier M.E. & Klisch Y. (2011) Learning and motivational impacts of a multimedia science game. *Computers & Education* **57**, 1425–1433.

Mioduser D., Tur-Kaspa H. & Leitner I. (2000) The learning value of computer-based instruction of early reading skills. *Journal of Computer Assisted Learning* **16**, 54–63.

Muratet M., Viallet F., Torguet P. & Jessel J. (2009) Une ingénierie pour jeux sérieux. (Engineering for serious games). *Proceeding of Conférence EIAH – Atelier Jeux Sérieux*, pp. 53–63. Le Mans, France.

Papastergiou M. (2009) Digital game-based learning in high school computer science education: impact on educational effectiveness and student motivation. *Computers & Education* **52**, 1–12.

Prensky M., ed. (2001) Digital natives digital immigrants. In *On the Horizon*, Vol. 9 pp. 1–6. MCB University Press.

Radillo A. (2009) L'expérimentation de l'utilisation des jeux vidéo en remédiation cognitive. (Experimentation on videogames use in cognitive remediation). *Enfances & PSY* **44**, 174–179.

Raybourn E.M. (2007) Applying simulation experience design methods to creating serious game-based adaptive training systems. *Interacting with Computers* **19**, 206–214.

Rebetez C. & Betrancourt M. (2007) Video game research in cognitive and educational sciences. *Cognition, Brain, Behavior* **11**, 131–142.

Robertson D. & Miller D. (2009) Learning gains from using games consoles in primary classrooms: a randomized controlled study. *Procedia – Social and Behavioral Sciences* **1**, 1641–1644.

Schittek M., Mattheos N., Lyon H.C. & Attström R. (2001) Computer assisted learning. A review. *European Journal of Dental Education* **5**, 93–100.

Schulz K.F., Altman D.G. & Moher D. (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Annals of Internal Medicine* **152**, 726–732.

Serruya M.D. & Kahana M.J. (2008) Techniques and devices to restore cognition. *Behavioural Brain Research* **192**, 149–165.

Sibbald B. & Roland M. (1998) Understanding controlled trials: why are randomised controlled trials important? *British Medical Journal* **316**, 201.

Sitzmann T. (2011) A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology* **64**, 489–528.

Squire K.D. (2008) Video game-based learning: an emerging paradigm for instruction. *Performance Improvement Quarterly* **21**, 7–36.

Sung Y.-T., Chang K.-E. & Huang J.-S. (2008) Improving children's reading comprehension and use of strategies through computer-based strategy training. *Computers in Human Behavior* **24**, 1552–1571.

Szilas N. & Sutter-Widmer D. (2009) Mieux comprendre la notion d'intégration entre l'apprentissage et le jeu. (Understand better the concept of integration between learning and game). *Proceeding of Atelier Jeux Sérieux: conception et usage, 4ème Conférence francophone sur les EIAH*. Le Mans, France.

Tanes Z. & Cemalcilar Z. (2010) Learning from SimCity: an empirical study of Turkish adolescents. *Journal of Adolescence* **33**, 731–739.

Tuzun H., Yilmaz-Soylu M., Karakus T., Inal Y. & Kizilkaya G. (2009) The effects of computer games on primary school students' achievement and motivation in geography learning. *Computers & Education* **52**, 68–77.

Vogel J.J., Vogel D.S., Cannon-Bowers J., Bowers C.A., Muse K. & Wright M. (2006) Computer gaming and interactive simulations for learning: a meta-analysis. *Journal of Educational Computing Research* **34**, 229–243.

Watson W.R., Mong C.J. & Harris C.A. (2011) A case study of the in-class use of a video game for teaching high school history. *Computers & Education* **56**, 466–474.

Westera W., Nadolski R., Hummel H. & Wopereis I. (2008) Serious games for higher education: a framework for reducing design complexity. *Journal of Computer Assisted Learning* **24**, 420–432.

Wrzesien M. & Raya M.A. (2010) Learning in serious virtual worlds: evaluation of learning effectiveness and appeal to students in the E-Junior project. *Computers & Education* **55**, 178–187.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1. Shared and specific criteria of initially selected serious games (SGs) and video games (VGs) and justification for their inclusion