



A critical review of the effectiveness of narrative-driven digital educational games

Jackson, Luke Conrad, O'Mara, Joanne, Moss, Julianne and Jackson, Alun C. 2018. A critical review of the effectiveness of narrative-driven digital educational games, *International journal of game-based learning*, vol. 8, no. 4, Oct-Dec, pp. 32-49.

DOI: [10.4018/IJGBL.2018100103](https://doi.org/10.4018/IJGBL.2018100103)

©2018, IGI Global

Reproduced with permission.

Downloaded from DRO:

<http://hdl.handle.net/10536/DRO/DU:30116165>

A Critical Review of the Effectiveness of Narrative-Driven Digital Educational Games

Luke Conrad Jackson, Deakin University, Geelong, Australia

Joanne O'Mara, Deakin University, Geelong, Australia

Julianne Moss, Deakin University, Geelong, Australia

Alun C. Jackson, Melbourne Graduate School of Education, University of Melbourne, Melbourne, Australia

ABSTRACT

Digital games are currently viewed, by many within the field of education, as a way to engage and motivate students, and to assist them in acquiring content knowledge and skills. Despite the growing interest in using digital games, including serious games, this is the first critical review of the literature on the effectiveness of digital narrative-driven educational games. Of 2550 articles initially screened, 130 were synthesised in the final review. The results suggest that such games can be more much more effective than traditional instruction for promoting attitude change, engagement, motivation, and skill acquisition; slightly more effective in promoting enjoyment and knowledge acquisition; and equal in fostering behaviour change. As such, they may be viewed as effective and versatile tools for teaching and learning.

KEYWORDS

E-Learning, Empathy, Media In Education, Multimedia/Hypermedia Systems, Narrative Immersion, Pedagogical Issues, Simulations, Student Engagement, Virtual Reality

INTRODUCTION

At the turn of the 20th Century, Sigmund Freud (1908) noted that “The child’s best-loved and most intense occupation is with his play or games” (p. 421). For Bettelheim (1987), play permitted the child “... to resolve in symbolic form unsolved problems of the past and to cope directly or symbolically with present concerns. It is also his most significant tool for preparing himself for the future and its tasks” (p. 170). Human beings play in structured and unstructured ways. Our earliest play is often unstructured, with items found at hand, with rules improvised to promote fun and social cohesion. In this way, we learn how the world works, and better understand our role in it (Brown, 2009). Structured games have existed for millennia, during which time they have taken on many guises, from the athletic games of the Ancient Greeks to board games, card games, children’s games, and – most recently – digital games.

Such games, once maligned within classrooms for their perceived potential to encourage violent behaviour (Anderson, 2004; Anderson & Bushman, 2002) and addiction (Griffiths, Davies, & Chappell, 2004), are now viewed by many in a more favourable light. James Paul Gee, one of the best known and most widely cited proponents digital educational game research, suggests that digital games, when well-constructed, offer the player/learner information on demand, just in time, and within a context that they care about (2003). Good games, he suggests, present problems in a logical

DOI: 10.4018/IJGBL.2018100103

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

order, and allow the player to develop increasingly complex skills through an ascending process of challenge and mastery, a process that he likens to the development of expertise in any domain. Gee (2011) has written, also, about the role of narrative in commercial and educational games. He argues that it is by walking in the shoes of the story's central character that the player can develop empathy as well as the ability to critically reflect upon their own behaviour in real life. In this process, such games have the potential to encourage the player to explore, engage in personal meaning-making and play with social boundaries (Shaffer, Squire, Halverson, & Gee, 2005).

Despite the rising number of studies that have sought to explore the effectiveness of digital educational games, a large-scale review has not yet been undertaken to synthesise what is known about the effectiveness of one of these games' most potentially powerful facets: narrative. This review seeks to address this gap in the literature.

Key Definitions

For the purposes of this review, a 'game' is defined as "... a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome" (Salen & Zimmerman, 2004, p. 83). A 'digital educational game' refers to a game produced for dissemination via a digital device for the purpose of educating the player. For a game to be considered narrative-driven, it must contain all six aspects of narrative as defined by Chatman (1980): events, character/s, setting/s, structure, point of view, and time. 'Effectiveness', within this critical review, refers to the various educational measures used by researchers as they attempt to discover whether a game has achieved its desired educational outcome, be that the acquisition of knowledge and/or skills, and/or attitude or behaviour change.

REVIEW OF LITERATURE

Many researchers and authors strongly believe in the potential of digital educational games to foster knowledge transfer, skill acquisition, and attitude or behaviour change, while others have argued for their effectiveness in term of engagement, motivation, and enjoyment over traditional curricula. However, gaining a consensus on their effectiveness is difficult. Examining meta-analyses, systematic and critical reviews of the scientific literature reveals a field in flux: publications are proliferating quickly (Boyle et al., 2016), yet there are suggestions of publication bias (Sitzmann, 2011). The application of different standards for study design and measurement continue to present problems for those who wish to compare cases (All, Castellar, & Van Looy, 2013; Ke, 2009; Tobias, Fletcher, Dai, & Wind, 2011; Young et al., 2012; Young, Slota, & Lai, 2012). These findings prompted All, Castellar, and Van Looy to design a set of guidelines which, they hope, will be used to standardise research on the effectiveness of digital games in future (All, Castellar, & Van Looy, 2016). Similarly, Clark, Tanner-Smith, & Killingsworth (2016) have called for game researchers to provide "thicker descriptions" of game-based interventions, to allow for more informed assessments to be made of the effect of various game elements.

Games and Narrative

Abdul Jabbar & Felicia (2015) investigated the link between engagement and two separate variables: role-play and narrative. A number of studies within their review suggested that role-play was an effective way of achieving immersion in the game, and that by playing the role of another person within the game students were better able to develop a sense of responsibility. Of the nine studies which addressed the use of narrative, the majority found that narratives helped students to relate to the situations and characters depicted in the game, which led to higher levels of understanding and the motivation to complete a greater number of missions. They also referred to the potential for narrative to grab and hold the player's attention, increasing their engagement with the learning materials and

making tasks seem less boring. Two studies within their review ran counter to these findings, by suggesting that narrative did not contribute the students' learning, or to their enjoyment of the games.

Clark, Tanner-Smith, & Killingsworth (2016) conducted a meta-analysis to investigate the effectiveness of digital games on learning outcomes in game versus non-game conditions and using augmented versus standard game designs. While they concluded that digital games significantly enhanced learning within both comparison groups, their assessment of the value of narrative within games was less positive. Their findings indicated that games with no story or thin story depth (only setting, scenery, or context) had significantly larger effects relative to those with medium story depth (involving some evolving story over the course of the game). However, they acknowledged that the lack of games with thick story depth in their study ($n=5$) made it difficult to reliably judge such games' effectiveness. Perhaps surprisingly, the researchers also discovered that there was a significant negative relationship between contextualisation (the use of immersive visual and narrative techniques) and learning outcomes. While these results may at first seem discouraging for those who wish to create more immersive games for learning, the researchers pointed out that no meta-analysis can account for all possible design approaches or qualities of implementation, and suggested that future research should not be limited to the highest-performing game characteristics identified within this study. They singled out visual and narrative contextualisation in this regard, but reiterated that their meta-analysis had identified some of the potential challenges for game designers.

To date, only one systematic review has been undertaken specifically investigating the effectiveness of using storylines within digital learning materials. Elena Novak (2015) located more than 70 articles, reports, dissertations, books, and conference presentations relating to the use of storylines, narrative, fantasy, or story structure in such materials, and deemed 11 to be suitable for synthesis. The studies included in the review were restricted to those containing both control and experimental groups, and were selected from a period of approximately 35 years, with four of the resources profiled being published between the early 1980s and the late 1990s. Within this relatively small sample, Novak found mixed results for storyline-related instructional effectiveness, with a number of studies suggesting that including a storyline either decreased or had no effect on learning, while one experimental study with elementary students showed positive learning outcomes.

METHOD

This critical review is different from the systematic and critical reviews aforementioned because it focuses entirely upon narrative-driven digital games, including primarily educational games, Serious Games, COTS games and Simulations used for educational purposes. These games can be aimed at any age, available via any delivery mode, in any genre, and which are designed to be played in any context. However, this review differs from its predecessors in two other distinct ways.

Firstly, while the researchers acknowledge the value of future studies adopting the sort of rigorous, scientific approach suggested by All, Nuñez Castellar and Van Looy (2016), this critical review has been designed to include quantitative, qualitative and mixed method studies so long as they have been peer-reviewed and report their findings clearly. The decision to adopt this approach was in direct response to the search strategy adopted by some previous researchers, who have been highly restrictive in the articles they would accept for inclusion. Because of this, they have been unable to consider a number of studies where the findings – even if less scientifically rigorous in their design or measurements – may nonetheless be of value for understanding why narrative-driven educational games have, or have not, achieved their educational aims.

Secondly, the decision has been made in the present review not to impose a definition of 'effectiveness' upon the studies, but to allow the term to be defined by the original designers of each study, in order to increase the sensitivity of the search for literature. In practical terms, this means that a study is deemed 'effective' if it has achieved all of its educational aims, whatever they may be. It is deemed to be 'negative' if none of these aims were achieved, or if – in the case of a study

using a control condition – the game condition actually learned less than the control. It is considered ‘mixed’ if some but not all of the educational aims were achieved.

While many large-scale reviews have considered studies which aim to achieve one educational outcome through the use of digital educational games (referred to, typically, as ‘knowledge acquisition’), this review takes in studies that measure attitude change, behaviour change, engagement, enjoyment, knowledge acquisition, motivation, skill acquisition, or ‘other’ (which usually refers to affective states such as flow or curiosity, as well as traits such as identity formation). The importance of this approach lies in the notion that reviews which consider only one of these dimensions are not only going to find fewer studies to synthesise, but are likely to brand a study as ineffective in achieving its educational aims when in fact it was partially effective. It is only by interrogating the data in this granular fashion that one can begin to appreciate which aspects of education games are most readily able to address, and which might remain a challenge.

A critical review methodology was selected to locate and analyse these studies because it is both flexible and rigorous, while allowing for critical judgment to be exercised by the author. As Grant and Booth (2009) explain:

A critical review provides an opportunity to ‘take stock’ and evaluate what is of value from the previous body of work. It may also attempt to resolve competing schools of thought. As such, it may provide a ‘launch pad’ for a new phase of conceptual development and subsequent ‘testing’. (p. 3)

In order to strengthen the findings, and allow for other studies to be reviewed over time, primary sources were located and synthesised based upon the principles of the Preferred Reporting Items for Critical Reviews and Meta-Analyses (PRISMA) Statement (Moher et al., 2009). This Statement aims to standardise and encourage best-practice in systematic and meta-reviews by providing reviewers with a clear and precise methodology for conducting reviews, including a checklist (Moher et al., 2009, pp. 876-877). The principles of this procedure were adhered to when literature was collected and reported on, with detailed descriptions provided on the search strategy used; inclusion and exclusion criteria; and data extraction methods, all of which are illustrated by a Prisma Flow Diagram (Moher et al., 2009, p. 877). The discussion is predominantly narrative in structure to allow for the explication and comparison of results.

Search Strategy

For the initial screening process, inclusion criteria included quantitative, qualitative, or mixed-method research; studies conducted in the developed world in any educational setting; and studies reporting the use of narrative-driven digital games for the purpose of education. Based upon these criteria, six education and technology-related databases were searched, using the search terms ((narrat* OR stor* OR fantasy) AND (digital OR computer OR video OR internet OR mobile OR app) AND (educat* OR serious OR applied) AND game AND effect* NOT toy NOT “digital storytelling”). Limiters included studies published in the English language, in scholarly (peer-reviewed) journals, from 2000-2016. Ultimately, this process elicited a total of 2,502 articles. Other sources of potentially relevant studies were reviewed for inclusion, including five of the top 25 highest-impact Education, Social Science, Computer Science, and Computer-human Interaction-related journals for 2014, as defined by SCImago Journal & Country Rank, a portal that includes the journals and country scientific indicators developed from the information contained in the Scopus® database (Elsevier B.V.). These publications, which had an h-index of between 35 and 93, were hand-searched. The combined total of articles located by this process was 2694. After this initial search phase was complete, all references were exported into an EndNote library, where 144 duplicates were eliminated. This yielded 2550 articles to be retained for screening using Covidence, a web-based software platform designed to streamline the production of systematic reviews.

Inclusion and Exclusion Criteria

The title and abstract of each article uploaded to Covidence was read to assess whether it met the criteria for inclusion. A paper was excluded if it reported on educational games that were not digital, such as paper-based or board games, or on game-like experiences that were not games, including simulations or transmedia campaigns that did not incorporate a game. Similarly, a study was excluded if it sought to measure the effectiveness of games in which only limited aspects of narrative were present, such as character and/or setting. A study was excluded if it attempted to measure the effectiveness not of the game as whole but of a particular mechanic (such as the use of ‘narrative agents’) or social dynamic (single-player versus multiplayer) on a particular game. In such studies, ‘effectiveness’ is often measured not against a control group who has not played the game, but against other players playing the same game under different conditions, meaning that the label of ‘effective’ is relative, and therefore relevant, only to that study.

After the titles and abstracts of 2550 references were screened, 2282 were deemed irrelevant, while 268 were deemed suitable for further examination. After examining the full text of these references, 130 were found to meet the inclusion criteria, while 138 were excluded with reasons recorded.

Data Extraction

The resulting data were grouped into a single table (can be found at the following webpage: http://www.lukecjackson.com/s/Review-of-Effectiveness_Appendix.docx), that was designed to outline the process by which a game’s effectiveness was tested. This table contains the following categories: Authors; Country of Origin; Game Purpose; Mode of delivery; Context for Play; Players per Game; Game Genre; Subject Area/s; and Overall Outcome. It also contains Educational Outcomes, including Attitude Change, Behaviour Change, Engagement, Enjoyment, Knowledge Acquisition, Motivation, Skill Acquisition, and Other; Participant age; Number of participants; Duration of Intervention; Method; and Results.

Data Analysis

This critical review was designed to answer the question, “How effective are narrative-driven, digital educational games in achieving their educational aims?” In order to fully address this central question, it has been necessary to investigate a number of sub-questions, which form the basis of the Results and Discussion section.

RESULTS AND DISCUSSION

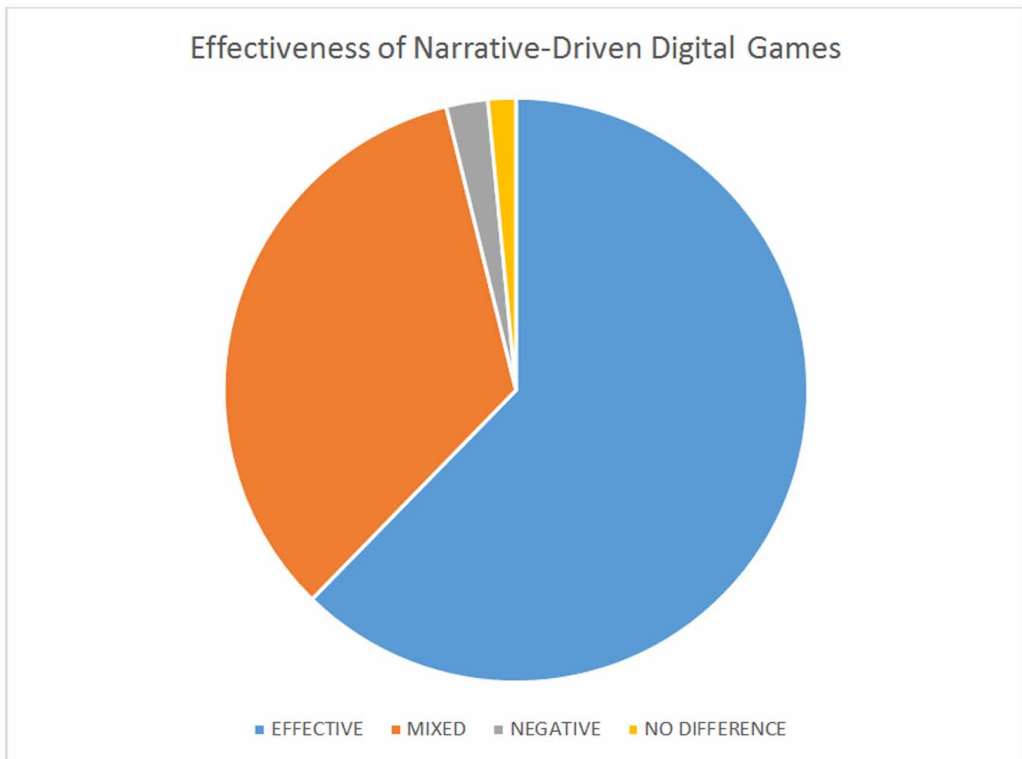
Understanding the Effectiveness of Game Play Elements

In order to appreciate the relative effectiveness of the various elements that may influence gameplay and/or learning, the 130 studies included in this critical review have been categorised by a range of features that apply to all games, and each of these has been considered individually. Where possible, brief descriptions of specific game studies which describe the use of such games have been included for illustrative purposes.

How Many Games Were Effective in Achieving Their Educational Aims?

The results of this critical review show, first and foremost, the high proportion of the 130 studies which have demonstrated positive or mixed results for the effectiveness of using narrative-driven, digital games in an attempt to achieve their educational aims (62.3% and 33.8% respectively); this is compared to only 2.3% of studies in which none of the study’s educational goals were achieved, and 1.5% of studies in which there was no difference between the results of students playing games and a control group. It should be noted that, in the majority of cases where a study achieved mixed results, the researchers did not find an improvement in one area, such as knowledge acquisition, but

Figure 1. Effectiveness of Narrative-Driven Digital Games



noted gains in other areas, particularly engagement, motivation, and enjoyment. The only exception was where there were inconsistent results within a single educational outcome when, for instance, the same study had been undertaken using two games, or two classes of students. The total proportion of studies listed under each category of effectiveness are shown in Figure 1.

What Type of Studies Have Been Conducted to Investigate the Effectiveness of These Games?

The methodological and demographic composition of the studies contained in this review represent a broad spectrum, from rigorous randomised control trials to small-scale interventions with qualitative analyses of student and teacher data. This reflects the concern, noted in previous reviews into the effectiveness of games, that there is a lack of homogeneity in the methods employed in measuring effectiveness (All et al., 2013; Boyle et al., 2012; T. M. Connolly et al., 2012; Jabbar & Felicia, 2015). The studies selected for synthesis were undertaken by researchers from 28 countries working alone or in partnership, with five countries accounting for most of the studies compiled for this review. These are the United States, with American researchers contributing to 51 (or 39.2%) of the total; Taiwan, who contributed to 16 studies (or 12.3% of the total); and Spain, the Netherlands, and Great Britain, who contributed to 9, 8, and 7 studies respectively. The dominance in this area of inquiry by the United States is in keeping with the findings of previous research into the broader area of effectiveness of digital games (Boyle et al., 2016). Interestingly, of the 130 studies of narrative-driven digital educational games covered in this review, 64 were undertaken by teams that included game designers, while only two acknowledged the inclusion of a dedicated writer.

Effectiveness by Design Purpose

One of the remarkable aspects about the games utilised for these studies of effectiveness was the high proportion of games whose primary focus is educational (94.6%) compared with Commercial Off-The-Shelf (COTS) games (4.6%). While COTS games made up a small section of the results, 4 of the 6 studies involving COTS games deemed these games to have been effective in achieving all of their educational aims, as was the one study to cover both COTS and Educational games. While these numbers are small, they seem to offer at least tentative support for previous claims that COTS games can be used effectively in an educational setting (Connolly, et al., 2012; Boyle et al., 2016).

Effectiveness by Subject Area

Reviewing effectiveness by subject area is something that researchers have called for in recent years (Wouters et al., 2013). The subject areas most often targeted by these interventions were Science (n=34), Mathematics (n=22), and Work Readiness (n=10). According to the selected articles, they were effective in 70.6%, 68.2% and 60% of cases respectively. A group of subjects associated with the Language Arts, including English Language, Literacy, Reading, Vocabulary, and Writing, were addressed by 20% of the studies selected for synthesis (n=26). They were deemed to have been effective in a combined 69.2% of cases. These findings, and others, are illustrated in Figure 2.

Effectiveness by Target Age Group

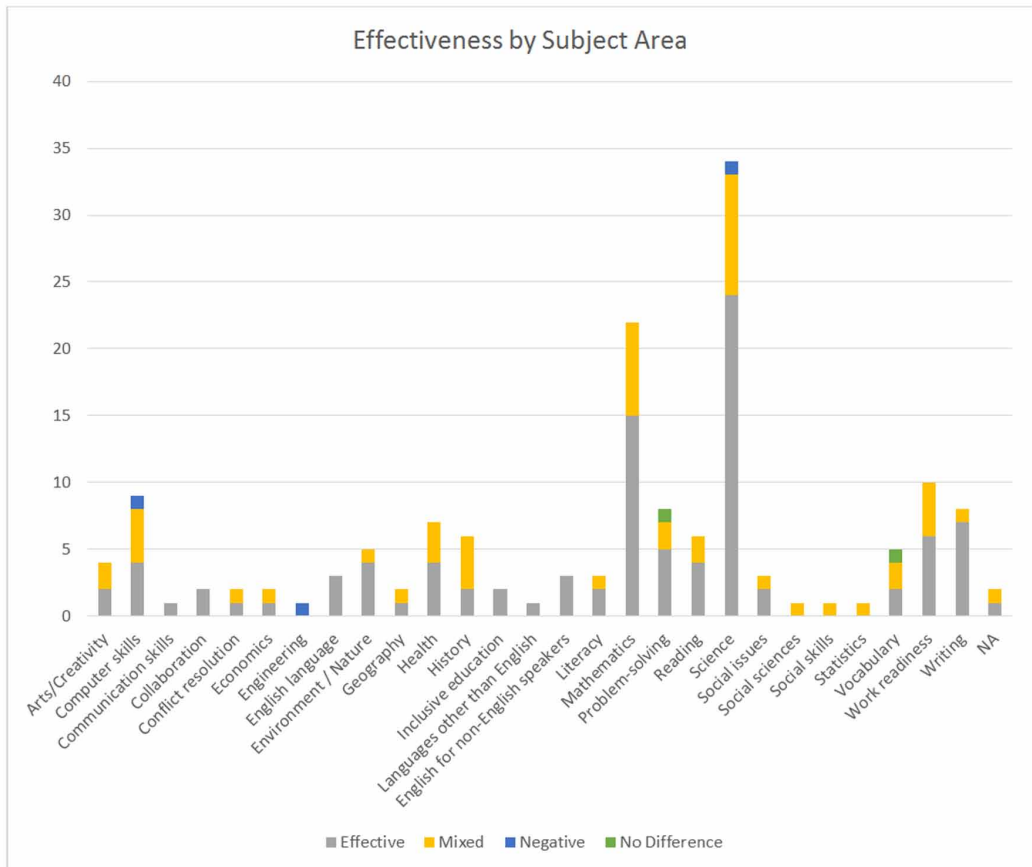
The largest 'target audiences' for these studies were players of secondary school age (n=48) and primary school age (n=38), followed by 'tertiary' and 'mixed ages' (n=20 and n=18 respectively). Of these four major cohorts, games were found to be least effective with tertiary participants, with only 45% of studies concluding that games had met all of their educational objectives. Within the other three cohorts, at least 50% of studies were judged to show that games had been effective for participants, with studies focusing on primary and secondary school-aged participants the most favourable (73.7% and 64.6% respectively).

Effectiveness by Genre

Previous research into the effectiveness of both educational and COTS games on students aged 14 or older has revealed a lack of variation in the game genres used, with most studies focusing on either simulations or puzzles. These authors have recommended, therefore, that games of different genres be tested in order to assess their level of success in achieving learning outcomes (T. M. Connolly et al., 2012). Within the current review, 31 genres (including some hybrids) and categories were identified. In many cases, the genre had been noted by the authors of the original study; where this was not the case, the most likely genre was assigned by the authors of this review, based upon the perceived alignment between the game (as described) and genre hallmarks. It should be noted that, rather than offering a definitive categorisation, game genres are often fluid, their boundaries blurred or even contested, making genre selection a fraught process (Apperley, 2006).

The game genre most heavily represented within this selection of studies was the Role-Playing Game, or RPG, with a total of 21.5%, made up of 'straight' RPGs plus RPGs with other elements such as mystery and strategy, as well as puzzle-based RPGs and Massively Multiplayer Online RPGs, or MMORPGs. Together, these were deemed to have been effective in a combined total of 71.4% of cases in which they were tested. Of these sub-categories, RPGs involving mystery (n=6) and Puzzle-RPGs (n=3) were deemed to have been 100% effective. The second most cited game genre was Puzzles (including puzzle-based hybrids), which accounted for 19.2% of the total number. They were also ranked in the middle of three most often cited genres in terms of effectiveness, with 64% of these games deemed entirely effective. Simulations, which tended to relate mostly to health or work preparedness, were the third most studied game genre within this cohort at 16.9% of the total, of which 54.5% were deemed to have been completely effective. These findings are illustrated in Figure 3.

Figure 2. Effectiveness by Subject Area



Effectiveness by Mode of Delivery

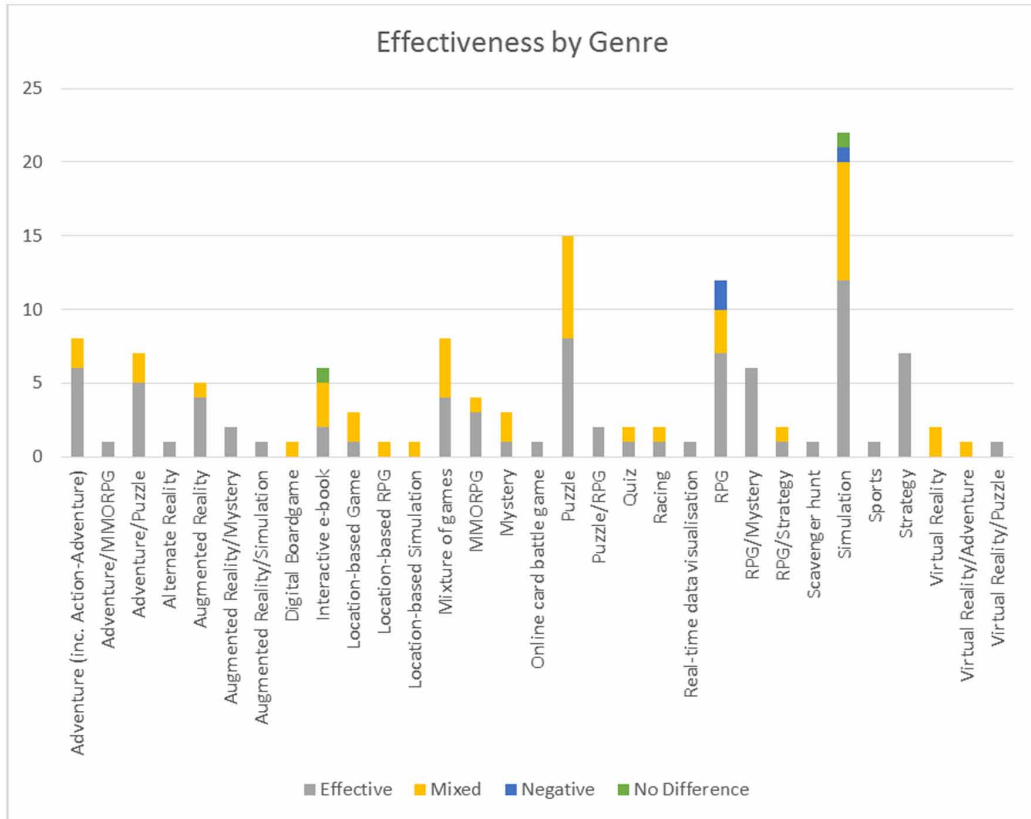
By far the most frequently used technology for delivering digital games was a computer, with 70.8% (n=92) of all games accessed in this way. Games accessed via computer may be played on a desktop or laptop, PC or Macintosh, and may be pre-loaded onto the system or network, or be accessed over the internet via a browser. Perhaps it is not surprising that, with such a high percentage of the games delivered in this mode, the percentage of games deemed to have been ‘effective’, ‘mixed’, ‘negative’, or to have made ‘no difference’ on computer (68.5%, 27.2%, 3.3%, and 1.1% respectively) closely correlates with the totals for all games.

The second most frequently used mode of delivery was mobile application (or ‘app’), which constituted 6.9% of the total (n=9), only 55.6% of which were deemed to have achieved all of their educational aims. All of these studies were published after 2010, which indicates that mobile applications are still an emerging trend in educational games research. Their success rate was lower than the related category of Augmented Reality Games (8.5% of the total), which were deemed effective in 63.6% of cases, with the remainder achieving mixed results.

Effectiveness by Context for Play

Perhaps not surprisingly, given the high proportion of games created specifically for educational purposes, 108 (or 83.1%) of the games reported on within these studies were primarily tested in a school or institutional setting, such as a classrooms, school library, computer laboratory, or a combination

Figure 3. Effectiveness by Genre



of locations. Of these games, 64.8% achieved all of their educational aims, 30.6% showed mixed results, 2.8% achieved none of their educational aims, and 1.9% were no different to the control group against which they were measured. The high proportion of games tested within educational settings is understandable, given the practical considerations associated with monitoring and measuring play that goes on outside the classroom. This may account for the relatively small number of studies in which students played entirely at home ($n=5$). Games designed to be played in a specific location, such as on a city street or in a meeting, made up 11.5% of the total ($n=15$), with the percentage deemed entirely effective in just 46.7% of cases, while the majority (53.3%) achieved mixed results.

Effectiveness by Players Per Game

The majority of games tested within the selected studies were single player games ($n=76$), followed by multiplayer games ($n=22$), then games designed to be played in groups of two or more ($n=21$), with a small number of games offering both single player and multiplayer ‘modes’ ($n=6$). Findings suggest that multiplayer games can be even more effective than single player ‘campaigns’ in achieving the desired educational outcomes (77.3% compared with 64.5%). While games aimed at groups must be played ‘on site’, so that players can be in close physical proximity, this is not necessarily the case for multiplayer games, many of which can be played via a network or the internet, making it possible for learners to play the game whether they are at an educational institution, at home, or - if they are rolled out to various industries – presumably at separate workstations on the job. However, some

caution must be exercised when viewing the results of these data. In many cases, the intended number of players for the game may have differed from the actual number of players, due to local factors which affected implementation. Similar issues were noted in other studies, where students might play the game independently, but then share information or strategise about how to solve in-game tasks outside of the game (Liao, Chen, Cheng, Chen, & Chan, 2011; Lim, 2008), making it difficult to ascertain exactly how much of the positive or negative influence of the game could be attributed to in-game, rather than environmental, factors.

Effectiveness by Educational Outcome/s

It was common for studies within this data set to measure more than one educational outcome. Because of this, it was possible for a study to show little or even negative change in one educational outcome while demonstrating significant gains (either as commonly referred to or statistically-speaking) in another. The 130 studies examined within this critical review set out to measure, quantitatively and/or qualitatively, a total of 287 educational outcomes, including 'Attitude Change', 'Behaviour Change', 'Engagement', 'Enjoyment', 'Knowledge Acquisition', 'Motivation', 'Skill Acquisition', and 'Other'. It is by understanding the measure of achievement on these learning outcomes that one can fully appreciate the potential for narrative-driven digital educational games in the classroom. Of the 287 educational outcomes measured within these studies, 224 were deemed to have been effective, 35 mixed, 7 negative, while there were 21 instances in which an educational outcome was deemed to have been equal to, but no more effective than, the results of a control group. These results are illustrated in Figure 4.

Given that the vast majority of games included in these studies were aligned with a set curriculum, it seems logical that the most often tested educational outcome was Knowledge Acquisition, which was tested in 113 (or 86.9%) of studies. Within these studies, the researchers' goals for Knowledge Acquisition were achieved in 79 (or 69.9%) of cases. The next most studied outcome was Engagement, which was examined in 43 (33.1%) of studies, and was deemed to be effective in 38 (88.4%) of cases. Motivation was the next most measured outcome, featuring in 34 (or 26.2%) of studies, where it was demonstrated to have been effective 88.2% of the time. This was closely followed by Skill Acquisition, which was measured in 33 (or 25.4%) of studies, and was shown to be the most effective of all educational outcomes at 90.9%. Despite the very high ratings for both Engagement and Motivation, 'Enjoyment' (which was measured in 14.6% or 19 studies) was only deemed to have been effective in 68.4% of cases. By contrast, while Attitude Change was only measured in 11.5% of studies (n=15), it had an effectiveness rating of 86.7%.

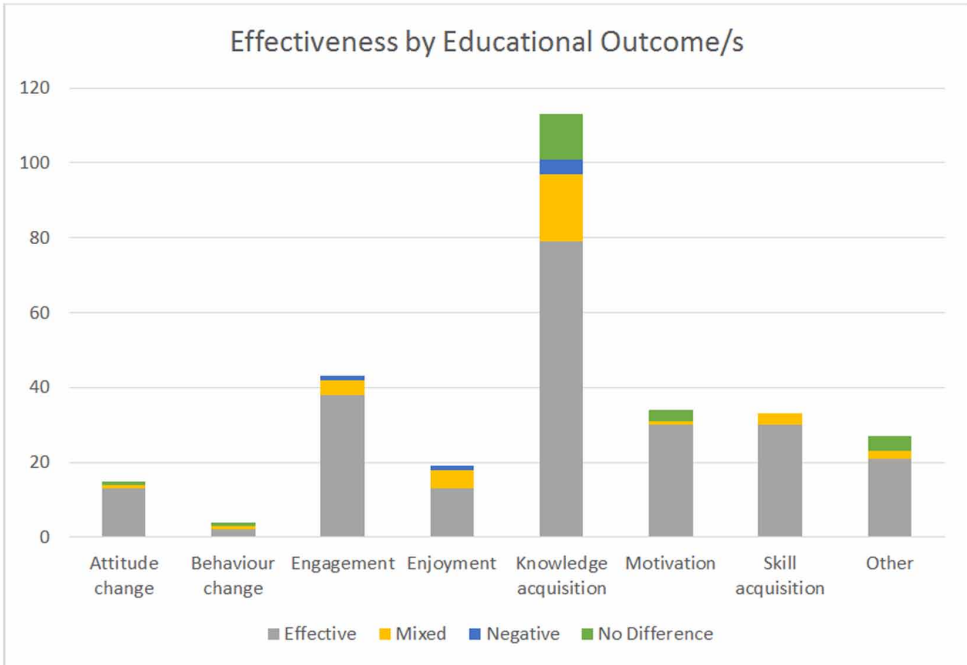
The least effective educational outcome, according to the studies in which it was measured (n=4), was Behaviour Change, which had just a 50% success rate. Gordon and Schirra (2011) used an RPG at a master planning meeting in Boston's Chinatown neighbourhood in an attempt to promote understanding of local issues and community engagement. They observed that, "While playing a character clearly made an impression on players ... it did not translate to an immediate, rational reframing of the issues within the meeting itself", and suggested that "role-play, within the magic circle, is quite effective in altering decisions that get made only within the magic circle. Outside, different rules apply" (pp. 184-185).

Limitations of the Review

While every attempt has made to be comprehensive in the location and synthesis of relevant research, the researchers do not claim that this review is exhaustive. It summarises the research on narrative-driven digital educational games based upon the search terms used, the databases included, and the time period of the review.

Each study included in this critical review has been reported on using its 'own terms' for assessing effectiveness. While the researchers believe that this has maintained the integrity of the studies' original findings, it is important to acknowledge that results can only be descriptive, since no meta-analysis of the includes studies' findings is possible.

Figure 4. Effectiveness by Educational Outcome/s



Although every effort was made, during each screening phase, to include all relevant interventions, if the study’s authors did not provide enough detail about the game to decide whether it met the inclusion criteria, the researchers removed the article rather than allowing its results to erroneously influence the findings. In order to attempt to include as many suitable interventions as possible, before dismissing a potentially relevant intervention, the researchers visited publicly-accessible web pages to assess whether the game was, indeed, both educational and narrative-driven. This approach has ensured that every study that made it into the synthesis reported on an intervention in which the effectiveness of a narrative-driven educational game was assessed.

Previous studies have noted that educational game- and simulation-related literature suffers from some publication bias (Sitzmann, 2011). Studies have been selected for review based upon stringent criteria previously outlined, but the researchers acknowledge that, if not for this bias, a higher proportion of less effective studies may have been included in a review, and the results may therefore have looked different.

CONCLUSION

The majority of studies included in this review suggest that narrative-driven, digital educational games are at least partially effective in most cases, with 62.3% of studies reporting that the game/s they tested met all of their intended educational outcomes, and a further 33.8% suggesting that games achieved at least some of their intended outcomes. The number of studies that reported negative findings (in which the game condition performed worse than a control group) was 2.3% of the total; while 1.5% of studies reported on games that showed no effect beyond that experienced by a control group. These findings are more positive than those of Ke (2009), who determined 52.3% of studies were effective, 26.2% were mixed, 1.5% were negative, and 18.5% showed no difference above the control

condition. The two biggest differences between the current review and Ke's qualitative meta-analysis are in the number of studies determined to be 'effective' and to have shown 'no difference' from the control. This is accounted for by the decision, in the current review, to accept, as 'effective', any study reporting positive results that exceed those of the control group, even if the level of achievement is not statistically significant, in order to allow for a wider range of studies, and results, to be evaluated.

Based upon the findings of this review, it appears that the effectiveness of narrative-driven digital educational games is not limited to games targeting a specific subject area, learning outcome, or target age group; to games made in a specific genre, for a particular mode of delivery or context for play; or to games involving a particular number of players. Nor does success seem to rely upon games being designed specifically for the purpose of education or – on the other hand – specifically for entertainment. In short, these findings demonstrate not only the educational effectiveness of narrative-driven, digital educational games, but their versatility as tools for teaching and learning. These assertions are in keeping with those of Vogel et. al (2006), who suggested that interactive games and/or simulations can lead not only to higher levels of knowledge acquisition, but to better attitudes towards learning, regardless of the gender or age of the player or whether players played the games on their own or with others.

The high level of effectiveness of games for motivation, at 88.2% of the 34 games in which it was tested, stands in stark contrast to the findings of Wouters, Van Nimwegen, Van Oostendorp, & Van Der Spek (2013), whose meta-analysis suggested that serious games were more effective for learning and retention, but less motivating, than conventional instructional methods. However, while some games within this review, such as those targeting attitude and behaviour change, might constitute 'serious games', many others would not, making the findings difficult to compare.

A key difference between this review and many of those that have preceded it is that, while there can be a temptation to suggest that a game is either effective or ineffective, almost one third of studies achieved mixed results. It is only by examining these results in detail that a more complete picture can be gained of what works, and what does not, in developing narrative-driven, digital educational games. These results demonstrate that, even when the researchers did not find the expected level of improvement in one area, such as Knowledge Acquisition, these were contradicted by other positive findings, particularly in the areas of Skill Acquisition, Engagement, and Motivation. In the case of location-based games, for example, while they were deemed to have been entirely effective just 46.7% of the time, closer examination reveals these games were highly engaging and motivating for players. In the studies in which they were tested, engagement and motivation achieved effectiveness ratings of 100% (6 of 6 studies) and 88.9% (8 of 9 studies) respectively.

Many of the most effective games referred to within the studies that form this critical review share some key features that can be informative for those who seek to design such games or to select suitable games for their own students to play. Firstly, players within these games are able to embody an authentic role, which assists them to view themselves, and each other, as more than just 'students' (Chee, 2010; Chee & Tan, 2012; Nash & Shaffer, 2011; Romero-Hall, Watson, Adcock, Bliss, & Adams Tufts, 2016). In an extension of this phenomenon, the role of 'player as tutor' has also been explored with success, as participants are required to offer their content-related expertise in order to assist (Marsh et al., 2011) or even nurture their in-game companions (Liao et al., 2011).

Secondly, many effective games allow for, and encourage, communication between players, either during play, as with multiplayer or team games, or in class-based reflection sessions afterwards. This interaction with other participants provides players with a way of consolidating subject knowledge and fostering skills such as problem-solving and critical thinking (Thomas M. Connolly, Stansfield, & Hainey, 2011; Shih, Shih, Shih, Su, & Chuang, 2010; Squire & Jan, 2007). These findings are consistent with those of Papastergiou (2009), whose critical review of the empirical evidence on the use of games in health and physical education revealed that such games have the potential to foster collaboration and teamwork both in the local and global contexts. In a study of one such game, *Alien Contact!* (Dunleavy, Dede, & Mitchell, 2009), researchers designed a "jigsaw pedagogy" in

which players worked in teams of four to investigate the reasons behind an alien landing. As they scoured a location with handheld devices at the ready, each player took on a different role (Chemist, Cryptologist, Computer Hacker, or FBI Agent), and received different pieces of the puzzle, which would only make sense when shared with their teammates. Not only was this engaging for students, but it led to a deeper appreciation of their teammates, as observed by their teacher:

They all took on an identity. They all felt strong ownership that they were an expert at this...and if they didn't have roles they may not have been as eager to work together because they really did need each person. (Dunleavy et al., 2009, p. 15)

In terms of the narrative elements that might be utilised to create effective games, the findings from this review are quite illuminating. Rieber (1996) called for educational games to employ 'endogenous' rather than 'exogenous' fantasy elements. This is supported by studies included in this critical review, which suggest that the drama must be intertwined with, and dependent upon, the learning tasks (Ke, 2008; ter Vrugte et al., 2015). When narrative, game mechanics and learning are completely balanced, the player may not even be aware that they are playing a game designed for learning (Marsh et al., 2011). As the creators and researchers of one such game, *Quest Atlantis*, have pointed out, "Possibly due to the immersive simulation context (perceptual immersion) or to the compelling nature of the story (narrative immersion), literally all students appeared to be engaged, with actual shouts of glee occurring when the teacher announced that it was time to work ..." (Barab, Sadler, Heiselt, Hickey, & Zuiker, 2007, p. 68).

It seems clear that games would be more likely to keep the player interested by placing the learning in a context which is meaningful and engaging (Barab, Pettyjohn, Gresalfi, Volk, & Solomou, 2012; Rubino, Barberis, Xhembulla, & Malnati, 2015). This has been achieved in a number of cases by adhering to the traditional 'Hero's Journey' (Dickey, 2006; Lopez-Arcos, Padilla-Zea, Paderewski, Gutierrez, & Abad-Arranz, 2014), while others have stressed the importance of making the dramatic stakes of the player's actions clear. Barab et al. (2012) suggest that these stakes have the potential not only to engage the player's interest, but to subvert some of the traditional expectations of what constitutes authentic learning within the contemporary classroom. After introducing Secondary School students to a game-based intervention with a focus on the ethics of medical research, they concluded that "... It might be argued that the fictional consequentiality afforded by the game play was more 'real' in terms of legitimizing the content and the student than was the real-world potential of doing better on a future test" (p. 531). Similarly, Tüzün, Yılmaz-Soylu, Karakuş, İnal, & Kızılkaya (2009) noted that fourth and fifth graders playing such a game, which was designed to teach them about Geography, demonstrated not only higher levels of knowledge acquisition, but achieved statistically significant increases in intrinsic motivation and decreases in extrinsic motivation after play, which translated into less focus on grades as the sole marker of success.

Finally, the findings of this review have reinforced the importance, for players, of having an emotional connection with the game, allowing them to develop a sense of empathy, which is the key to fostering critical reflection (Gee, 2011) as well as a range of other learning outcomes, including knowledge and skill acquisition, engagement, enjoyment, motivation, immersion and identification. All of these have been shown to increase when the player has someone to care about, whether that someone is the point-of-view character (Chee, Tan, & Liu, 2010; Lopez-Arcos et al., 2014; Nash & Shaffer, 2011); another player's avatar in a multiplayer game (San Chee & Tan, 2012; Zheng, Young, Wagner, & Brewer, 2009); another person whom the player can talk to, in the real world, face-to-face during gameplay (Facer et al., 2004; Gordon & Schirra, 2011); or a suitably sympathetic Non-Playing Character (Christopoulos, Mavridis, Andreadis, & Karigiannis, 2011; Romero-Hall et al., 2016; Tice et al., 2009).

Suggestions for Further Research

The researchers would join with All, Castellar, and Van Looy (2013) in recommending that 'best practices' be adopted for the design and reporting of future research into the effectiveness of DGBL.

In the case of studies reporting on the use of narrative-driven games, we would recommend that ‘best practice’ should allow researchers to isolate particular narrative elements, such as character or setting, in order to measure their effect upon learning outcomes. In order to adequately investigate the effect of individual elements, it may be necessary to look beyond a single dimension of learning, such as ‘knowledge acquisition’, and investigate the way that a particular narrative element, such as a ‘character’, affects the player’s performance in relation to a range of educational outcomes.

Very few of the games synthesised for this review report on behaviour change as a measure of effectiveness (n=4). Of those instances where behaviour change was measured it was found to be effective in two cases, mixed in one case, and to have made no difference in the other. With such small numbers, it is difficult to make any definitive statement about what whether narrative-driven digital educational games are an effective approach to teach behaviour change. This is an area that should be explored further. Given the inherent difficulties noted by Gordon and Schirra (2011) between translating attitude change within a game to behaviour change outside the ‘magic circle’, it would seem that such games – if they were to be developed - should focus on behaviour change not in isolation, but as the final step in a continuum. This continuum, it would seem, should move the player from a state of engagement and enjoyment, through knowledge and skill acquisition, to motivation and attitude change, before ideally arriving at the desired change in behaviour. Narrative elements should work to reinforce the educational aims of the game by engaging the player’s interest, effectively structuring their experience, and allowing them to empathise with others as they take on new roles.

REFERENCES

- Abdul Jabbar, A. I., & Felicia, P. (2015). Gameplay Engagement and Learning in Game-Based Learning: A Systematic Review. *Review of Educational Research*, 85(4), 740–779.
- All, A., Castellar, E. P. N., & Van Looy, J. (2013). A systematic literature review of methodology used to measure effectiveness in digital game-based learning. *Paper presented at the European Conference on Games Based Learning*.
- All, A., Castellar, E. P. N., & Van Looy, J. (2016). Assessing the effectiveness of digital game-based learning: Best practices. *Computers & Education*, 92, 90–103.
- Anderson, C. A. (2004). An update on the effects of playing violent video games. *Journal of Adolescence*, 27(1), 113–122.
- Anderson, C. A., & Bushman, B. J. (2002). The effects of media violence on society. *Science*, 295(5564), 2377–2379.
- Apperley, T. H. (2006). Genre and game studies: Toward a critical approach to video game genres. *Simulation & Gaming*, 37(1), 6–23.
- Barab, S., Pettyjohn, P., Gresalfi, M., Volk, C., & Solomou, M. (2012). Game-Based Curriculum and Transformational Play: Designing to Meaningfully Positioning Person, Content, and Context. *Computers & Education*, 58(1), 518–533.
- Barab, S., Sadler, T., Heiselt, C., Hickey, D., & Zuiker, S. (2007). Relating Narrative, Inquiry, and Inscriptions: Supporting Consequential Play. *Journal of Science Education and Technology*, 16(1), 59–82. doi:10.1007/s10956-006-9033-3
- Bettelheim, B. (1987). *A good enough parent: A book on child-rearing*. Alfred A. Knopf.
- Boyle, E. A., Connolly, T. M., Hainey, T., & Boyle, J. M. (2012). Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior*, 28(3), 771–780. doi:10.1016/j.chb.2011.11.020
- Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., & Pereira, J. et al. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, 94, 178–192. doi:10.1016/j.compedu.2015.11.003
- Brown, S. L. (2009). *Play: How it shapes the brain, opens the imagination, and invigorates the soul*. Penguin.
- Chatman, S. B. (1980). *Story and discourse: Narrative structure in fiction and film*. Cornell University Press.
- Chee, Y. S. (2010). Game-Based Learning as Performance: The Case of Legends of Alkhimia. In *Proceedings of the European Conference on Games Based Learning* (pp. 47-54).
- Chee, Y. S., Tan, E. M., & Liu, Q. (2010). Statecraft X: Enacting citizenship education using a mobile learning game played on Apple iPhones. *Paper presented at the 2010 6th IEEE International Conference on Wireless, Mobile and Ubiquitous Technologies in Education (WMUTE)*.
- Chee, Y. S., & Tan, K. C. D. (2012). Becoming Chemists through Game-Based Inquiry Learning: The Case of “Legends of Alkhimia.” *Electronic Journal of e-Learning*, 10(2), 185-198.
- Christopoulos, D., Mavridis, P., Andreadis, A., & Karigiannis, J. N. (2011). *Using Virtual Environments to Tell the Story: “The Battle of Thermopylae.”*
- Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Review of Educational Research*, 86(1), 79–122.
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2), 661–686. doi:10.1016/j.compedu.2012.03.004
- Connolly, T. M., Stansfield, M., & Hainey, T. (2011). An alternate reality game for language learning: ARGuing for multilingual motivation. *Computers & Education*, 57(1), 1389–1415. doi:10.1016/j.compedu.2011.01.009

- Dickey, M. D. (2006). Game design narrative for learning: Appropriating adventure game design narrative devices and techniques for the design of interactive learning environments. *Educational Technology Research and Development*, 54(3), 245–263.
- Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and Limitations of Immersive Participatory Augmented Reality Simulations for Teaching and Learning. *Journal of Science Education and Technology*, 18(1), 7–22.
- Facer, K., Joiner, R., Stanton, D., Reid, J., Hull, R., & Kirk, D. (2004). Savannah: Mobile gaming and learning? *Journal of Computer Assisted Learning*, 20(6), 399–409. doi:10.1111/j.1365-2729.2004.00105.x
- Freud, S. (1908). Creative writers and day-dreaming. *Standard edition*, 9, 141–153.
- Gee, J. P. (2011). Stories, probes, and games. *Narrative Inquiry*, 21(2), 353–357.
- Gordon, E., & Schirra, S. (2011). *Playing with empathy: digital role-playing games in public meetings*. Brisbane, Australia.
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, 26(2), 91–108.
- Griffiths, M. D., Davies, M. N., & Chappell, D. (2004). Online computer gaming: A comparison of adolescent and adult gamers. *Journal of Adolescence*, 27(1), 87–96.
- Jabbar, A. I. A., & Felicia, P. (2015). Gameplay Engagement and Learning in Game-Based Learning A Systematic Review. *Review of Educational Research*, 85(4), 740–779.
- Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay? *Computers & Education*, 51(4), 1609–1620.
- Ke, F. (2009). A qualitative meta-analysis of computer games as learning tools. In *Handbook of research on effective electronic gaming in education* (Vol. 1, pp. 1–32).
- Liao, C. C. Y., Chen, Z. H., Cheng, H. N. H., Chen, F. C., & Chan, T. W. (2011). My-Mini-Pet: A handheld pet-nurturing game to engage students in arithmetic practices. *Journal of Computer Assisted Learning*, 27(1), 76–89. doi:10.1111/j.1365-2729.2010.00367.x
- Lim, C. P. (2008). Global citizenship education, school curriculum and games: Learning Mathematics, English and Science as a global citizen. *Computers & Education*, 51(3), 1073–1093.
- Lopez-Arcos, J. R., Padilla-Zea, N., Paderewski, P., Gutierrez, F. L., & Abad-Arranz, A. (2014). *Designing stories for educational video games: A Player-Centered approach*. Albacete, Spain.
- Marsh, T., Xuejin, C., Nickole, L. Z., Osterweil, S., Klopfer, E., & Haas, J. (2011). *Fun and learning: the power of narrative*.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–269.
- Nash, P., & Shaffer, D. W. (2011). Mentor modeling: The internalization of modeled professional thinking in an epistemic game. *Journal of Computer Assisted Learning*, 27(2), 173–189. doi:10.1111/j.1365-2729.2010.00385.x
- Novak, E. (2015). A Critical Review of Digital Storyline-Enhanced Learning. *Educational Technology Research and Development*, 63(3), 431–453.
- Papastergiou, M. (2009). Exploring the potential of computer and video games for health and physical education: A literature review. *Computers & Education*, 53(3), 603–622. doi:10.1016/j.compedu.2009.04.001
- Rieber, L. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development*, 44(2), 43–58.
- Romero-Hall, E., Watson, G. S., Adcock, A., Bliss, J., & Adams Tufts, K. (2016). Simulated environments with animated agents: Effects on visual attention, emotion, performance, and perception. *Journal of Computer Assisted Learning*. doi:10.1111/jcal.12138

- Rubino, I., Barberis, C., Xhembulla, J., & Malnati, G. (2015). Integrating a Location-Based Mobile Game in the Museum Visit. *J. Comput. Cult. Herit.*, 8(3), 1–18. doi:10.1145/2724723
- Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. MIT press.
- San Chee, Y., & Tan, K. C. D. (2012). Becoming Chemists through Game-based Inquiry Learning: The Case of Legends of Alkhimia. *Electronic Journal of e-Learning*, 10(2).
- Shaffer, D. W., Squire, K. R., Halverson, R., & Gee, J. P. (2005). Video games and the future of learning. *Phi Delta Kappan*, 87(2), 104–111.
- Shih, J.-L., Shih, B.-J., Shih, C.-C., Su, H.-Y., & Chuang, C.-W. (2010). The influence of collaboration styles to children's cognitive performance in digital problem-solving game "William Adventure": A comparative case study. *Computers & Education*, 55(3), 982–993. doi:10.1016/j.compedu.2010.04.009
- Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology*, 64(2), 489–528.
- Squire, K. D., & Jan, M. (2007). Mad City Mystery: Developing Scientific Argumentation Skills with a Place-based Augmented Reality Game on Handheld Computers. *Journal of Science Education and Technology*, 16(1), 5–29.
- ter Vrugte, J., de Jong, T., Wouters, P., Vandercruysse, S., Elen, J., & van Oostendorp, H. (2015). When a game supports prevocational math education but integrated reflection does not. *Journal of Computer Assisted Learning*, 31(5), 462–480. doi:10.1111/jcal.12104
- Tice, E., Tregubov, T., Schnipper, K., Park, Y.-K., diCiaccio, R., Friedman, M., & Loeb, L. (2009). GreenLite Dartmouth: unplug or the polar bear gets it.
- Tobias, S., & Fletcher, J. (2012). Reflections on "a review of trends in serious gaming.". *Review of Educational Research*, 82(2), 233–237.
- Tobias, S., Fletcher, J., Dai, D. Y., & Wind, A. P. (2011). Review of research on computer games. *Computer games and instruction*, 127, 222.
- Tüzün, H., Yılmaz-Soylu, M., Karakuş, T., İnal, Y., & Kızılkaya, G. (2009). The effects of computer games on primary school students' achievement and motivation in geography learning. *Computers & Education*, 52(1), 68–77. doi:10.1016/j.compedu.2008.06.008
- 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(3), 229–243.
- Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105(2), 249.
- Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., & Yukhymenko, M. et al. (2012). Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. *Review of Educational Research*, 82(1), 61–89.
- Young, M. F., Slota, S. T., & Lai, B. (2012). Comments on "reflections on 'a review of trends in serious gaming.'". *Review of Educational Research*, 82(3), 296–299.
- Zheng, D., Young, M. F., Wagner, M. M., & Brewer, R. A. (2009). Negotiation for Action: English Language Learning in Game-Based Virtual Worlds. *Modern Language Journal*, 93(4), 489–511. doi:10.1111/j.1540-4781.2009.00927.x

Luke Conrad Jackson is a writer of novels, games and other media, a teacher, and a researcher. His research explores the nexus of narrative, education, and technology.

Joanne O'Mara is located in Language, Literature and Literacy Education at Deakin University, Melbourne, Australia. Her research and scholarship focuses on emergent literacies and new textual practices; digital games; drama pedagogy; and the spatial, social and temporal dimensions of teachers' work.

Alun C. Jackson is an Honorary Professorial Fellow, Melbourne Graduate School of Education, University of Melbourne, Victoria, Australia; Honorary Professor, Faculty of Health, Deakin University, Victoria, Australia; Honorary Professor, Centre on Behavioural Health, University of Hong Kong, Hong Kong SAR, China; and Director, Australian Centre for Heart Health, Melbourne, Victoria, Australia.