Digital game-based L2 learning outcomes for primary through high-school students: A systematic literature review

Acquah, Emmanuel O.; Katz, Heidi T.

Published in:
Computers and Education

DOI:
10.1016/j.compedu.2019.103667
10.1016/j.compedu.2019.103667

Published: 01/01/2020

Document Version
Accepted author manuscript

Document License
CC BY-NC-ND

Digital game-based L2 learning outcomes: A systematic literature review

Abstract

This systematic literature review examines digital game-based L2 learning outcomes between 2014 and 2018, with a focus on participants 6 – 18 years old. The initial search yielded 567 results, from which a total of 26 articles were included in the final content analysis. From the analysis, it is evident that digital learning games (DLG) can benefit players' language learning, psychological state, and contemporary competences. In addition, there are key game features highlighted throughout the research that point to positive outcomes: challenge, player control (autonomy), goal orientation, collaboration, and ease-of-use. This paper also presents descriptives about research context, target language, gaming platform, research methods, intent of game, and game genre. Based on the findings, it is clear that DLGs are an effective tool, but future research should explore how they can best be implemented in the classroom.

Keywords

Digital learning games; L2 learning; systematic review; game-based learning; learning motivation

1. Introduction

Since the popularization of digital games, researchers in a wide range of fields have debated about their benefits and drawbacks. Parents are concerned by their overuse and the exposure to violence, while educators are worried about students' attention spans. However, there is growing evidence that in the right context, digital games can be used to enhance motivation and learning outcomes (Peterson, 2010; Squire, 2002; Squire, 2008).

Previous reviews on digital game-based learning (DGBL) have looked at positive affect in specific subject areas (Chiu, Kao, & Reynolds, 2012; Hung, Yang, Hwang, Chu, & Wang, 2018; Li & Tsai, 2013; Papastergiou, 2009; Peterson, 2016; Young, Slota, Cutter, Jalette, Mullin, Lai et al., 2012), the experience of flow or engagement (Perttula, 2017), trends in research (Hwang, 2012), and research methods used (All, 2014).

This review will utilize aspects of a framework created by (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012) in their systematic review examining literature concerning the positive impact of computer games and serious games in general, with later updates and adjustments in 2016 (Boyle et al., 2016; Hainey, Connolly, Boyle, Wilson, & Razak, 2016). Hung et al., (2018) provided a scoping overview of the empirical evidence on the use and impacts of digital games in language earning from 2007-2016. This review will expand on their review by analyzing the context of game-play and focusing on primary through high school children.

The purpose of this review is to better understand how digital games can be used to enhance language learning for students, especially within classroom settings. The aim is to provide an overview of the empirical evidence on the impacts of digital game-based language learning (DGBLL) on primary through high school age students between 2014 and 2018. The years were selected in order to overlap with the review of Hung et al., (2018) since aspects in this review were not discussed in theirs, and because different databases will be searched. The search will be conducted in order to answer the following research questions:

1. What empirical evidence is there concerning the positive impacts and outcomes of digital-game based language learning on primary through high school-age children?
2. Were there any significant associations between game characteristics, context of game-play, and impact on players?

1.2. Defining terms

Digital games used for the purpose of learning are discussed using a variety of terms: digital education game (DEG), serious game, digital game-based learning (DGBL), digital game-based language learning (DGBLL), game for learning, and digital learning game (DLG). While the various terms all emphasize a similar purpose - learning through the use of digital games - it is

important to define what constitutes an educational digital game, and how they will be referred to in this review.

This review will use the same terminology proposed by Hainey et al., (2016) in their systematic review for the game type based on developer intention: *game for learning* (GL) and *entertainment game* (EG). They defined EGs as "pre-made, COTS games that are used in [school] for the purposes of learning, teaching a particular subject or promoting engagement," while, "games for learning...is the production of a specially implemented application for the purposes of learning, teaching a particular subject of promoting engagement," (Hainey et al., 2016, p. 203). In their search, there were only a few games referred to as serious games, and so those fell under the category of GL (Hainey et al., 2016). In other words, the GL category includes serious games and games that have been developed for an educational purpose, and the EG category is comprised of commercial games that were developed for entertainment, but are also being used for educational purposes. When discussing these games as a whole, they will be called *digital learning games* (DLGs). As Klopfer, Osterweil, & Salen (2009) explain, "Learning Games may be associated with formal educational environments (schools and universities online or off), places of informal learning (e.g. museums), or self-learners interested [in] acquiring new knowledge or understanding," (p. 21). In this study, the focus will be on digital games used within both formal and informal educational environments, "that target the acquisition of [language] as its own end and foster habits of mind and understanding that are generally useful or useful within an academic context" (Klopfer et al., 2009).

1.3. Digital learning games

The popularization of DLGs stems from a change in learners. Students in the 21st century grow up with modern technology (televisions, computers, iPad, and smartphones) and thus are digital natives. This exposure to technology impacts what kind of learners enter our schools. Oblinger (2004) refers to these students, born in or after 1982, as the Net Generation (NetGen'ers). NetGen'ers tend to prefer teamwork, experiential activities, the use of technology, and they expect learning to be fun (Oblinger, 2004; Prensky, 2001). These new kinds of learners are more self-directed and thus their learning styles align more with the constructivist approach that is often found in digital-games (Oblinger, 2004): games activate prior knowledge and then scaffold learning and provide instant feedback; games are typically situated, meaning learning occurs in a particular context and that knowledge can later be transferred to real-life; users get to learn through experience, problem-solving, and failure; and digital games are often social environments (Shaffer, Squire, Halverson & Gee, 2005; Hamari, Shernoff, Rowe, Coller, Asbell-Clarke & Edwards, 2016; Klopfer et al., 2009; Oblinger, 2004; Prensky, 2001; Squire, 2008). It is not surprising that students are drawn to the freedom they experience while playing digital games, as opposed to the traditional schooling environment where teachers struggle to keep students engaged.

Throughout history, philosophers and theorists - including Plato, Vygotsky, Piaget, and Rousseau - have expressed the positive role of play in a child's development and education, and DLGs are a modern tool to efficiently integrate play into 21st-century classrooms (Wilkinson, 2016). Children participate in play for its intrinsic purposes, and DLGs can connect that intrinsic motivation to learning due to the game's ability to differentiate and give students a sense of autonomy (Hamari et al., 2016; Peterson, 2010; Squire, 2008; Wilkinson, 2016).

1.4. Language Learning

Several reviews have examined the impact of DGBLL (Chiu et al., 2012; Hung et al., 2018; Peterson, 2016). Research shows that out of all school subject areas, DGBL tends to be most often used, and associated with positive outcomes in language learning (Godwin-Jones, 2018; Wastiau, Kearney, & Van den Berghe, 2009; Yip & Kwan, 2006; Young et al., 2012). This could be due to the fact that language learning is more effective when situated and when learners can be social; two components found in DGBL (Peterson, 2016). As previous reviewers mentioned though, the existing literature related to DGBLL is scarce (Hung et al., 2018; Peterson, 2016), and thus this review will add to the existing literature from a new angle. Hung et al., (2018) searched 10 journals for research on DGBLL at all ages, including L1 learning, and using different platforms. This review deviates from theirs by:

1. looking at primary through secondary school age participants
2. excluding L1 learners
3. Analyzing the context of game-play (e.g. classroom, home, etc.)
4. expanding the databases searched (Hung et al., (2018) used the Web of Science database)
5. using different search terms

2. Method

2.1. Protocol registration

The protocol for this review was registered as "Systematic Literature Review on Digital Learning Games" in the Open Science Framework (OSF) on November 29, 2018 (https://osf.io/2ncgj/). There have been several adjustments made since the original protocol; these changes, along with explanations, will be described throughout the paper.

2.2. Inclusion criteria

The following criteria were used to determine the papers included in the review:

1. Game type: a) games for learning/serious games or b) commercial computer games/entertainment games that are used for educational purposes
2. Game platform: tablet, computer or mobile device
3. Can be an online game
4. Peer-reviewed articles

5. Articles published from 2014 through November 2018.
6. Written in English
7. Studies on children from primary school through high school (age six and up)
8. Discusses language learning (L2: second-language or foreign language) from playing a DLG
9. Discusses outcomes related to language learning or motivation.

In addition, the following is a list of exclusion criteria:

1. Board games, simulations (unless simulation is the game genre), digital tools
2. Game platform: video console
3. Systematic reviews, meta-analyses, book chapters, position papers
4. L1 learning
5. Related to learning by game design

Originally, mobile devices were under the exclusion criteria as they aren't as commonly used as a tool in a classroom and many mobile language applications are not games. However, during the screening process, this decision was reconsidered since many mobile games can also be played on tablets. Language applications that were not in a gaming format were still excluded, but it was decided that mobile games could be included. There were also two studies where the game console was unknown, but these were included with reason: Shahriarpour & kafi (2014) used L.A. Noire, a game that could be played on a video game console, but also on a computer; Sundqvist & Wikstrom (2015) researched informal out-of-school gameplay so it was likely that at least some game-play occurred on a computer, tablet, or mobile phone. In addition, some studies included participants whose age range exceeded our age criteria, but they were included as long as the overlap between the range and our age criteria was significant (e.g. 4-12, 13-31, and 12-34). Finally, in the screening there were several articles on digital games used for bilingual participants' learning one of their two languages, but these articles were excluded.

2.3. Databases searched

The electronic databases searched were selected based on their pertinence to education, information technology, linguistics, and psychology research: EBSCO (Academic Search Premier), ERIC (Education Resources Information Center), Emerald, Web of Science, Wiley Online, Taylor & Francis Online, and SAGE Journals.

2.4. Search terms

The search terms used were related to DLGs, outcomes (learning and motivational), and language acquisition. The search included similar terms to those used by Connolly et al., (2012), and additional terms to reflect this review's focus on language. These terms were chosen, rather than ones used by Hung et al, (2018) for two reasons: they specify only the use of digital games, and to include broader terms that could expand the possible types of outcomes from DGBLL. In

addition, while games played on video consoles are in the exclusion criteria, the term “video games” was used in the search as it is a broad term used to describe all types of digital games.

Based on the interest in digital games that stimulate learning, the following terms were used to narrow the search on games:

> (“computer games” OR “video games” OR “serious games” OR “digital game-based learning” OR “digital game-based language learning” OR “digital learning games” OR “digital education games” OR “digital language games” OR “digital games” OR MMORPGS OR MMOG OR MUD OR “online games”)

Next, the following terms related to language were used:

> (“language acquisition" OR "language learning" OR "second-language" OR "foreign language" OR "language education" OR "language class" OR “L2”)

Finally, to look for outcomes these terms were entered:

> (impacts OR effects OR affect OR evaluation OR outcomes OR engagement OR motivation OR skills)

Each search set was separated by the *AND* operator.

2.5. Study selection

Two researchers participated in the entire screening process. After screening for duplicates, the researchers met to discuss the criteria and spent time screening the first twenty titles and abstracts together. The articles were then split up between the researchers in order to complete the initial screening. This phase of the screening process was straightforward as the researchers were mainly focused on whether there was a mention of language learning, the age of participants, the device games were played on, and also the document type. If some of these factors weren’t mentioned in an article’s abstract, the authors still included the article in order to search the full paper for that information during the next screening. Any relevant meta-analyses and systematic reviews were placed in a separate folder. The researchers returned to these articles after the first screening and looked at the reference section. This allowed articles to be added snowballing.

In the second screening phase, the researchers met again to assess the aim, methods, results, and conclusion of each paper. The researchers tested for inter-rater reliability with ten papers, finding they were in agreement 90% of the time based on Cohen’s Kappa (1960). This screening was an iterative process, where the researchers worked independently but met to discuss any uncertainty. They also checked again for inter-rater reliability with the last eight papers and were in agreement 87% of the time.

A large majority of research had to be excluded due to the participants being University students, with the assumption that participants in those studies may have been chosen out of convenience.

Hopefully, future researchers will consider looking at younger participants as it is common for L2 learning to begin in primary school.

The PRISMA flow diagram (Fig. 1) was used to present the flow of information throughout the review process. The PRISMA statement was developed in order to provide a systematic way for authors to conduct and report on systematic reviews and meta-analyses, particularly in the field of health research (Moher, Liberati, Tetzlaff, & Altman, 2009; Moher, Shamseer, Clarke, Ghersi, Liberati, Petticrew, Shekelle, & Stewart, 2015). While we were unable to register our review with Prospero (ineligible due to not being related to health), per their recommendation, we did follow the PRISMA statement guidelines to complete our review (see Moher et al., 2015), and as stated previously, the protocol was uploaded to OSF instead.

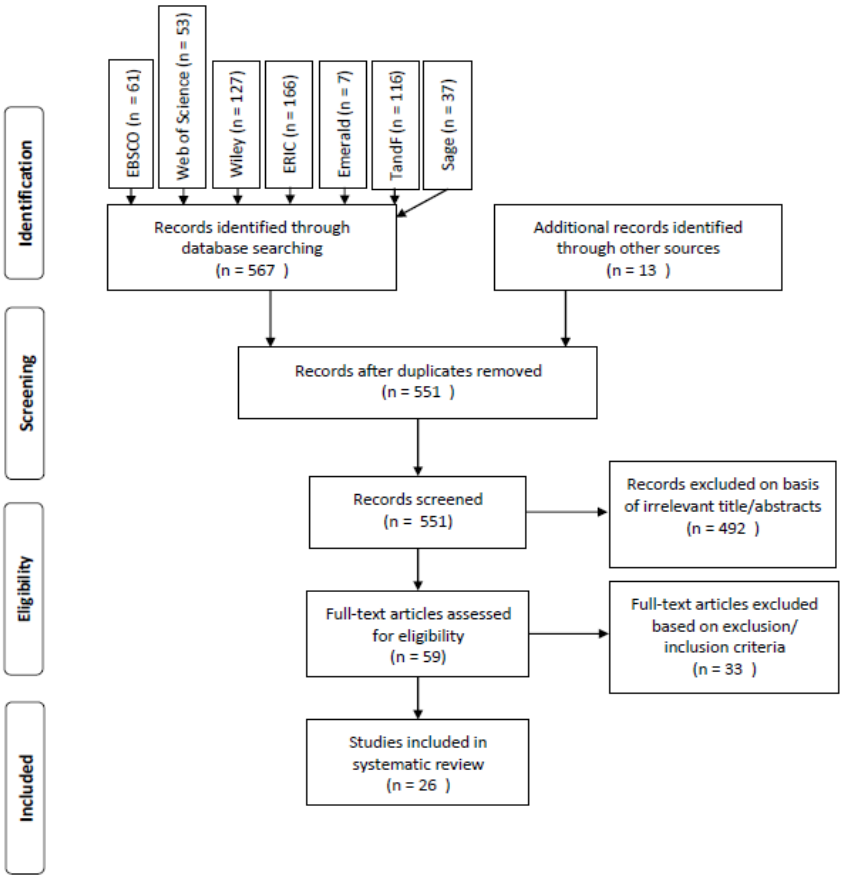

Figure 1. Flow diagram of the screening and selection procedure.

2.6. Data analysis

The two researchers conferred on how to classify the papers, and then independently coded them. Once again, problems in coding were discussed and the researchers shared and reviewed articles when necessary. The researchers used content analysis to examine the coded data, and descriptives from that analysis will be presented.

3. Results

3.1. Overview of included papers

As seen in Figure 1, a total of 26 articles were included in this review, only seven of which were also in the review conducted by Hung et al., (2018). Information about each article, including the authors, date, aims, methods, game details, and outcomes, are presented in the Appendix. There are three separate tables, with articles grouped by research design (experimental, quasi-experimental, and other). Ten of the studies used experimental research (38%), eight used quasi-experimental (31%), four used action research or design-based research (15%), two used correlational (8%), and just one used a case study and one an observational study (4%).

Figure 2 shows the number of studies using each design, based on the number of participants in the study. The largest study was correlational, with 3,945 participants (age 4-12), followed by an experimental study with 241 participants (age 12-18). The smallest study was classified as quasi-experimental (though this was inferred and not stated by the author), and had just four participants (Pennala, Richardson, Ylinen, Lyytinen, & Martin, 2014)

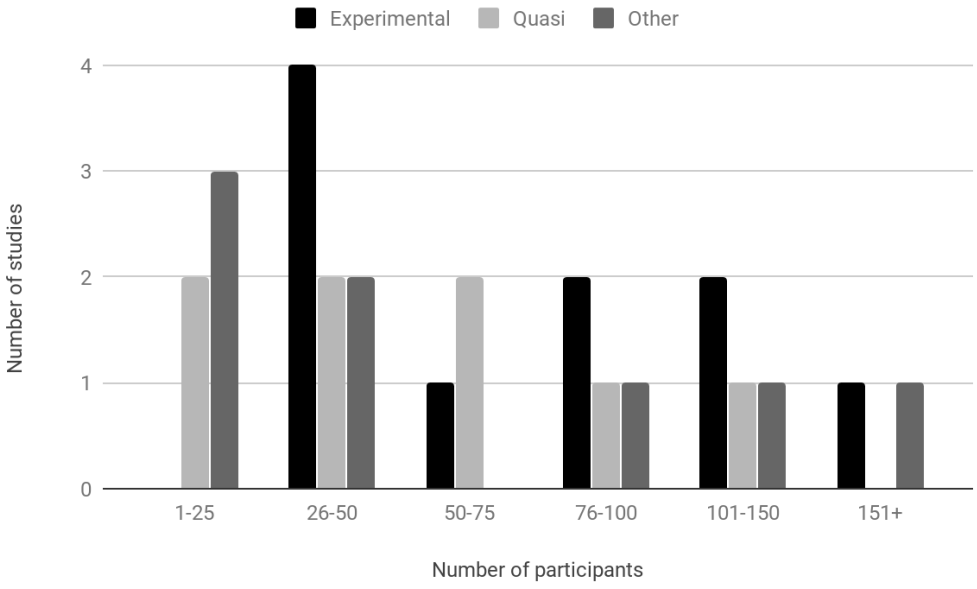

Figure 2. Number of participants by study design.

3.1.1. Research methods

In terms of types of data collected, the majority of studies used mixed methods (n = 14; 54%), followed by quantitative data (n = 11; 42%), and only one qualitative study (4%). However, coding the papers based on methods, as well as research design, posed a challenge. In many instances, the methods or research design weren't mentioned, which meant these classifications had to be inferred based on other information in the article. Also, some authors seemed to wrongly identify the research type as experimental when there wasn't randomization, but as a rule, the classification in the article was used, if available.

3.1.2. Language

Out of all the papers, English was the most common target language (n = 22; 85%), followed by Chinese (n = 2; 8%), then Finnish and Irish (n = 1; 4%). The two articles that researched Chinese DGBLL focused on teaching students Chinese radicals.

3.1.3. Game platform

Concerning game platform, computers were most often used (n = 20; 77%), followed by mobile phones (n = 3; 12%), unknown devices (n = 2; 8%), and tablets were only reported being used once (1%). It was a surprise that tablets and mobile phones weren't used more frequently, but it is possible that articles using tablets more often refer to them as "applications," a word that was purposefully excluded from our search, as discussed in section *2.2.* Two of the studies that used these platforms in their study explain they are beneficial because they can be used anywhere, anytime (Hwang, Shih, Ma, Shadiev, & Chen, 2016; Sandberg, Maris, & Hoogendoorn, 2014).

3.1.4. Country of research

An unexpected finding was where most of the research was conducted: eleven of the studies were conducted in East Asia (42%), seven from the Middle East (27%), seven from various regions in Europe (excluding Turkey; 27%), and one from Southeast Asia (4%). More specifically, nine studies were conducted in Taiwan (35%), and six in Iran (23%); it would be interesting to explore the reasons behind this research trend in these countries.

3.2. Game genre and features

Herz's (1997) classification was used in order to help code game genres, since it was often left unclassified in the articles. After the first attempt to code the articles by game genre, it was decided that educational mini-games and educational games would be added. This was due to the fact that the studies using educational mini-games often included a variety of games for learning with different game genres, and the educational games typically had limited information about the structure of the game or pictures revealed a very basic gaming environment that focused on learning. An example of this is from the description of Graphogame:

> The main idea of the game is that a player, listening through headphones, is required to respond to prototypical spoken items administered one at a time….Two or more visual items at a time are presented on the screen. The player has to choose the item that matches the auditory stimulus by clicking a mouse button...the game has many visual elements and graphics in common with simple computer games where the idea is to score as many points as possible (Pennala et al., 2014, p. 149-150).

None of the games used in the research articles were action, puzzle, fighting or sports. In addition, Sundqvist & Wikstrom (2015) studied out-of-school game-play, which means the genre of game-play, and intention of the game (GL or EG) varied per person and thus is unknown. The majority of the games used in these studies were either educational mini-games (n = 8; 31%); or educational games (23%). It's important to note, that one of the seven EGs (17 of the games were for learning, one unknown) were classified as mini-games. Figure 3 and Figure 4 present the number of games per genre and whether researchers were involved in either developing the game or reworking a pre-existing GL (commonly seen with simulation games). Figure 3 is showing only the GLs and figure 4 is displaying the EGs.

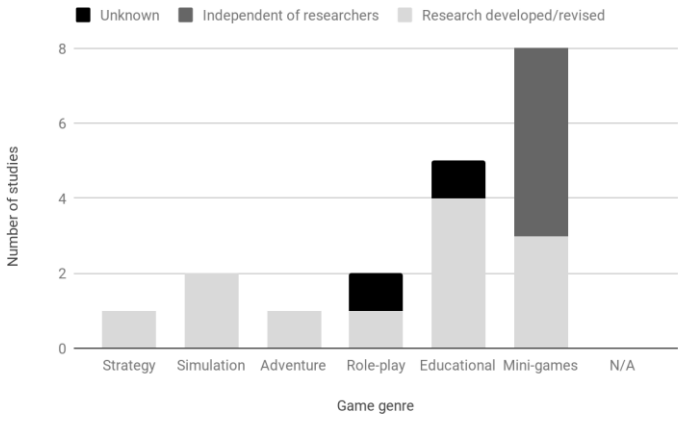

Figure 3. Games for learning: Number of studies in each game genre and whether researchers developed or revised the games.

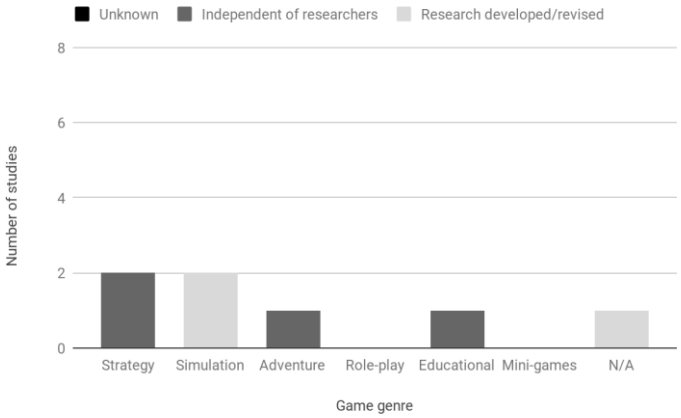

Figure 4. Entertainment games: Number of studies in each game genre and whether researchers developed or revised the games.

There were no clear patterns between game genre and outcomes, and not enough studies in each genre to draw conclusions, but more specific game features were observed to see whether they impacted players' motivation and learning. One study found the badge mechanism, which provides instant feedback, positively impacted players' self-efficacy and learning performance (Yang, Quadir, & Chen, 2016). A different study looked at which mini-games children chose to play most frequently and noted the features they shared: challenge, mystery, control, and having multiple players (Butler, Someya, & Fukuhara). Autonomy (player control) - was supported in another study by revealing a positive correlation with vocabulary learning, while challenge had a negative correlation with vocabulary learning as it can sometimes cause anxiety (Ebrahimzadeh

& Alavi, 2016). Therefore, it is important that games provide enough challenge, but players can still be successful. Challenge is also important when considering the structure and format of the game; ease-of-use must be considered, as learning how to play the game can be overwhelming and act as a deterrent if there are too many rules (Hwang et al., 2016; Wu, Chen, & Huang, 2014). Two more features were noted: Dourda, Bratitsis, Griva, & Papadopoulou (2014) found that 10 out of 17 participants reported collaboration as one of the best aspects of the DGBLL, while the main theme drawn from a study by Dalton (2016) was the importance of goal orientation. Thus, in designing a DLG, or using a DLG in the classroom, educators and designers should consider these features.

3.3. Outcomes from game-play

Outcomes from game-play were categorized based on the review by Hung et al., (2018), which identified six potential outcome categories: *language acquisition, knowledge acquisition, contemporary competencies, affective/psychological states, participatory behaviors,* and *correlational outcomes*. In this review, there were ten different types of outcomes related to *language acquisition* found in the articles: vocabulary, listening, reading, spelling, phonological awareness, orthographic processing, speaking, pronunciation, grammar, and unknown (the author did not specify). Only one article reported on *knowledge acquisition* (subject-matter, cultural or other types of knowledge gains) in geography (Dourda et al., 2014). Articles that reported on *contemporary competencies* included outcomes related to critical thinking, creative thinking, collaboration, communication, conversation, and problem-solving. Anything related to perceptions or attitudes, motivation, flow or engagement, self-efficacy or anxiety, cognitive load, and autonomy fell under the category of *affective/psychological states*. Participatory behaviors were more challenging to categorize, as only one study gave participants the option to stop playing the game. Thus, in this review, *participatory behaviors* are related to attrition, interaction between players, and impact on classroom interactions. The final category is correlational outcomes. *Correlational outcomes* are studies that looked at the relationship between individual characteristics and game-play or outcomes: gender, age, learning style, language learning anxiety, and gamers vs. non-gamers.

Quantitative or mixed methods articles were marked as having either *positive* outcomes or *negative* outcomes if the author reported significance in the results, with positive outcomes meaning DGBLL positively impacted the player. Articles with qualitative data were marked as *positive* or *negative* if the outcome was stated by the author as a positive or negative conclusive finding evidenced by the data. Anything coded as *no significant impac*t means the authors reported no significant findings or differences between groups (quantitative/mixed methods) or that findings were ambiguous and not completely supported, rather the authors stated their thoughts and suggestions for further research (qualitative).

Some articles focused on just one outcome from a given category, while others discussed multiple outcomes. For example, under learning acquisition one article reported vocabulary, listening, writing, and reading outcomes, some being positive and some being negative. Rather than reporting on each individual sub-category, all the outcomes under each of the categories were combined, meaning the aforementioned example would be coded as *mixed*. *Mixed* results could be a combination of positive and negative, positive and no significant impact, negative and no significant impact, or all three. Table 1 shows the number of positive, negative, no significant impact, and mixed results in each category. Correlational was excluded from the table since having a significant difference between genders, for example, wouldn't fit under the definitions of positive, negative, no significant impact, or mixed. However, there were six studies that reported correlational results.

Table 1
Outcome categories and the results

Outcomes	Positive	Negative	No significant impact	Mixed	Total
Language Acquisition	16	0	4	6	26
Knowledge Acquisition	1	0	0	0	1
Contemporary Competences	7	0	1	0	8
Affective/Psychological states	13	0	1	2	16
Participatory Behaviors	8	1	2	2	13
Total	45	1	8	10	64

Including correlational outcomes, there were a total of 70 reported outcome categories in the 26 studies. Some studies only looked at one outcome category, while others looked at a range of outcomes, but every study included some form of measurement for language acquisition. The most commonly analyzed sub-category for language acquisition was vocabulary.

In comparing the results for each outcome, 70% of the 64 outcomes (excluding correlational) were found to be positive. There was only one exclusively negative result, reported by Sandberg et al., (2014), whose study looked at attrition. However, other negative results were found, as seen by the 10 (16%) mixed results, but those studies also reported positive results or no significant impact. For example, Salehi (2016) found no difference in initial vocabulary learning between groups after the experiment, but found that the experimental group (used DLGs) had higher vocabulary retention one month later. Another study reported three different results from language acquisition: there was no significant difference between the experimental group (used DLGs) and control group on the immediate post-test for vocabulary; the control group performed better on the vocabulary retention test; and the experimental group performed better on the pronunciation test (Young et al., 2012). Several articles concluded that the studies needed to be

extended in order to draw solid conclusions on the games' impact (Hsu, 2017; Mazaji & Tabatabaei, 2016; Reitz, Sohny, & Lochmann, 2016).

Three studies looked at gender differences, with two finding significant differences between genders. In the three instances, it was clear that boys were more exposed to digital games than girls and thus tended to do better in terms of related outcomes (Mazaji & Tabatabaei, 2016; Sundqvist & Wikstrom, 2015). Butler et al., (2014) found differences in gender just in a couple of instances (depending on the game and level), but gender did not correlate with the frequency of game-play. However, they did find a correlation between age and frequency of game-play: the frequency of plays decreased as age increased. Other correlations observed in studies were gamers performing better than non-gamers (Sundqvist & Wikstrom, 2015); players with higher levels of English anxiety performing more complex learning and gaming behaviors, and having greater learning outcomes (Hwang, Hsu, Lai, & Hsueh, 2017); and two studies looked at learning styles, finding certain learning styles (serial learning style and sensing learning style) correlate with learning acquisition outcomes (Hsu, 2017; Wong & Hsu, 2016).

These results provide a clear picture of how DGBLL has been studied, and the types of outcomes it yields. Based on these findings, it is clear that DGBLL can be used effectively for language acquisition, contemporary competencies, and affective/psychological states. Given only one study looked at other knowledge acquisition, it is uncertain whether DGBLL can be used to enhance other knowledge as well. In addition, participatory behaviors are mostly positive, but given the challenge of coding them, it cannot be confirmed that DGBLL enhances participatory behaviors. Rather, most participatory behaviors overlapped with contemporary competences, since both looked at interaction and communication. Instead, it is suggested that future research looks into participants' choice to play, attrition rate, and how playing games impact the participants' activity in class as well. More detailed descriptions of the findings are provided in the Appendix.

3.4. Study context

Another area of interest was where the study took place since it is important to know whether educators can use DGBLL in their classrooms. There were three categories that articles could be coded into, though two studies took place in a research context that was never clearly identified, thus they were categorized as *unknown*. Since all 26 studies looked at language acquisition, Table 2 reveals the language acquisition result in relation to the environment the study was conducted.

Twelve studies (46%) were conducted in a *formal learning environment with educator facilitation* (FLF), which means they took place in a school or educational environment, and either a teacher or the researcher facilitated learning. This could be as simple as providing language guidance and support throughout the gaming experience or actually teaching lessons

along with the games. The majority of these studies yielded positive language outcomes (n = 8, 67%), with two outcomes being mixed and two resulting in no significant impact.

The *formal learning environment without educator facilitation* (FL) category means the study took place in an educational setting, but there wasn't any educator facilitation; the researcher or teacher may have given directions on how to play the game but wasn't involved in the facilitation of learning. Six studies (23%) took place in the FL environment, all of which resulted in positive language acquisition outcomes.

Four studies (15%) were conducted in an *informal environment* (IF), which includes at-home use or a laboratory-type context. Out of the four studies, one yielded positive results, one had no significant impact, and two had mixed results.

Two experimental studies were found to take place both in an FLF and an IF (8%). One study was facilitated within the classroom, but participants were also free to use the game for two weeks outside of the classroom, which resulted in positive speaking outcomes, but no significant difference in listening skills (Hwang et al., 2016). The other study also gave the participants the option to use the game outside of school after the initial instruction, and while there were positive grammar outcomes, the study also had a high attrition rate (Cornillie, Van Den Noortgate, Van den Branden, & Desmet, 2018).

Table 2
Study context and the results

	Formal learning environment with educator facilitation	Formal learning environment without educator facilitation	Informal environment	Formal with facilitation AND Informal	Unknown
Positive	8	6	1	1	0
Negative	0	0	0	0	0
No significant impact	2	0	1	0	1
Mixed	2	0	2	1	1
Total	12	6	4	2	2

4. Discussion

When researching the benefits of DLGs on L2 learning, it is important not just to look at outcomes related to language skills, but also outcomes that can enhance one's learning and learning experience. One key benefit of DLGs is the safe environment they provide for learners, which is revealed by learners' willingness to interact through the games (Reitz et al., 2016; Wu et al., 2014; Young & Wang, 2014). Language learning is effective when situated and requires the learner's motivation, confidence, and willingness to interact, but this isn't always easy to

accomplish in a traditional classroom (Zuengler & Miller, 2006). As shown by the positive results in this review, especially in relation to language acquisition, contemporary competencies and affective/psychological states, DLGs can be used as effective tools to use to enhance L2 skills and the overall learning experience. Previous research on DLGs suggests that the play and differentiation that occurs in DLGs can help activate player's intrinsic motivation and feelings of autonomy (Hamari et al., 2016; Peterson, 2010; Squire, 2008; Wilkinson, 2016), both of which were supported in this review.

In addition to finding evidence that supports the use of DGBLL, it was important to look at the context of game-play. Previous researchers have emphasized the importance of the teacher's facilitation of the game in order for it to be most effective (Beavis, Rowan, Dezuanni, McGillivray, O'Mara, Prestridge et al., 2014; Squire, 2002). As Beavis et al. (2014) explain, "What a games based learning environment in school actually becomes is closely tied to the way teachers think about games including what they believe can or cannot be achieved with games and how they believe games should or should not be used," (p. 570). This review found that even without teacher facilitation, DGBLL could be effective. However, it would be interesting for future researchers to conduct an experiment using the same game in multiple contexts and with various types of facilitation in order to see how DLGs can best be implemented within classrooms. The same concept could be applied to different gaming platforms, to examine whether one produces better gaming outcomes. In this review, it was not clear whether one gaming platform was better for learning, though the benefit of tablets and mobile phones is that they can be used anywhere.

When considering the use of digital games as learning tools, the features of the games are salient. Key features were discussed in several articles, mainly by themes that arose in interviews, though this too should be researched more. From the articles used in this review it seems that players need an easy-to-use game, that is challenging but enables success, provides instant feedback, gives the learner autonomy but also has a clearly defined goal, and allows for collaboration (Butler et al., 2014; Dalton, 2016; Dourda et al., 2014; Ebrahimzadeh & Alavi, 2016; Hwang et al., 2016; Wu et al., 2014; Yang, Quadir, & Chen, 2016). Of course, various learners have different needs and are drawn to different game features, but the value of DLGs is the fact that they can differentiate in many ways.

There are two key limitations in this study. Despite the evidence in favor of DGBLL, publication bias must be taken into consideration. As discussed previously, only one outcome category had a solely negative result, and even that one result was alongside positive results in other outcome categories. Thus, there were no published studies that revealed only negative results. Since this article only included peer-reviewed journal articles, it is possible that there are unpublished studies that yielded negative results. Second, while there were three several longitudinal studies that looked at retention, results were mixed and the maximum length of time was two months (Salehi, 2016; Shokri & Abdolmanafi-Rokni, 2014; Young & Wang, 2014). It would be important for future research to look at longer periods of time and whether DGBLL can be used consistently in a classroom to produce better results than other tools. Nevertheless, this review

shows that DLGs have been used to enhance language learning, in a variety of contexts, game genres, and with a range of participants. Further research into what makes certain games more effective than others should be explored.

5. Conclusion

This review discussed 26 studies looking into the impact of digital games on language learning and related outcomes from 2014 through November 2018. It was found that DGBLL can be used as an effective language learning tool that motivates players to learn and interact. DGBLL can be a fun, engaging, and challenging way to learn, and games allow for differentiation and learner autonomy.

Future research should seek to understand how games can be best implemented in classrooms so that teachers can use DLGs as tools. While this review found no difference between the gaming contexts and outcomes, more research into the topic would provide teachers with a better understanding of how to use the games, and their role in the facilitation. Wastiau et al., (2009) reported that, while the majority of teachers are curious about how to use games within schools, many don't because they are difficult to integrate into the curriculum and they lack time; teachers must take time to find games that relate to the curriculum, and then actually learn how to play them, which is an obstacle to integration. Therefore, teachers need to be given explicit guidance on how they can best utilize DLGs, without feeling like it is an additional burden. Digital devices are part of the 21st-century child's everyday life and can be used either as a tool or a toy. DLGs allow them to act as both, and it would benefit educators to know how to successfully use them.

Appendix

Below are three tables with information about each article. The tables are divided by study design. Some information under "Methods" and "Game details" has been abbreviated. A key with the abbreviations and formatting of the methods and game details section can be viewed after each table.

Table A1
Experimental Studies

Article	Aims	Methods* / Game details**	Outcomes
(Cornillie, Van Den Noortgate, Van den Branden, & Desmet, 2018)	To analyze L2 grammar practice with CALL materials in real contexts (outside of the lab) after one month of practice.	N= 126 (16-18); FLF & IL; English; Quan; Mixed factorial in vivo; web-based behaviour-tracking technology GL, Comp, MG: N/A, (researchers developed the games for the study)	Multilevel statistical analyses of accuracy and response time suggest that practice helped to develop automaticity, and that rule complexity and metalinguistic feedback played a role. The attrition rate was high.

(Ebrahimzadeh & Alavi, 2016)	To examine e-learning enjoyment as a predictor of DGB vocabulary learning and differences between watchers and players of a game.	N= 136 (12-18); FLF; English; Mixed for triangulation; e-learning enjoyment scale (EGameFlow), a vocabulary posttest, researcher field-notes EG, Comp, St: Warcraft III: The Frozen Throne - Defense of the Ancients	There was no significant difference between Players and Watchers, though both scores improved significantly ($p = .000$, partial $n^2 = .774$). Four dimensions had significant contributions to the total explained variance in vocabulary learning: challenge (-.350), immersion (-.291), autonomy (.326), and knowledge improvement (.388). In addition, e-learning enjoyment was found to significantly predict variance in vocabulary learning.
(Ebrahimzadeh, 2017)	To compare vocabulary acquisition between a commercial digital game (playing and watching) and traditional paper-and-pencil learning.	N= 241 (12-18); FLF; English; Mixed; pre- and post-vocabulary tests and researcher observations EG, Comp, St: Warcraft III: The Frozen Throne - Defense of the Ancients	Readers performed significantly lower than both players and watchers ($p = .001$), and there was no significant difference between players and watchers (.470).
(Hsu, 2017)	To develop and compare two Augmented Reality (AR) educational game systems for third graders to learn English vocabulary in free and situated surroundings.	N= 38 (3rd grade, M = 9); FL; English; Quan; pre-test of knowledge, post-test of vocabulary comprehension, and questionnaires of learning style, flow state, foreign language anxiety, and cognitive load. GL, Tablet, Sim: N/A (developed by researchers)	There was no significant difference between students using task-based (M = 79.91) or self-directed (M = 77.75) AR games, but both had high learning effectiveness. There was also no significant effects of learning style on students' flow state. However, the flow state of the self-directed group (M = 3.36) was significantly higher than the task-based group (M = 2.87), but flow was not correlated with learning anxiety or mental effort. In addition, students with a serial learning style had lower mental effort and anxiety.
(Hwang & Wang, 2016)	To investigate the students' performance and behaviors of learning English vocabulary with different guiding strategies.	N = 50 (6th grade); FL; English; Mixed; pre- and post-tests of knowledge (vocab), cognitive load questionnaire, behavior tracking with sequential analysis, and in-depth interviews. GL, Comp, RPG: N/A (developed by researchers using RPG maker XP)	Students learning with cloze guiding strategy (M = 88.91) had a significantly better learning performance with higher cognitive load than those with multiple-choice guiding strategy (M = 83.33). The game with cloze item guidance also engaged students in single and double-loop situated learning, while those using multiple-choice guidance only performed single-loop situated learning.
(Hwang, W., Shih, Ma, Shadiev, & Chen, 2016)	To test mobile game-based learning in enhancing English learning, as well as student perceptions and motivation in a situational game-based learning context.	N = 50 (senior high school); FLF & IL; English; Mixed; pre- and post test for listening and speaking, questionnaire on motivation and perceptions, and one-on-one semi-structured interviews. GL, Mobile, Ed: N/A (developed by researchers using Apache, PHP and MYSQL)	The experimental group (game-based; M = 10.40) significantly outperformed the control group (traditional teaching; M = 7.60) on the verbal post-test ($p = .009$), but there was no significant difference between groups on listening post test ($p = .377$). Student motivation was high and most students had positive perceptions toward the mobile system.
(Salehi, 2016)	To investigate the impact of Instructional Video Games (IVGs) on the vocabulary retention of Iranian language learners.	N = 85 (13-31); unknown context, possibly FL; English; Quan; vocabulary pre- and post-tests. GL, Comp, MG: gamestolearnenglish.com, "big describer," "hang man," "compare," etc.	There was no significant difference between groups in the post-test ($p=.13$). However, in the delayed post- test there was a significant difference ($p=.03$), with the experimental group (used IVG) having better retention in learning vocabulary.
	To investigate the effectiveness of using	N = 40 (14-15); FLF; English; Quan; spelling pre- and post-tests	There was a significant difference between groups in recall of spelling, with the

(Shokri & Abdolmanafi-Rokni, 2014)	computer games on recall and retention of word spelling in Iranian EFL male and female students.	GL, Comp, MG: Hangman, Spelling bee, Fast hand, and Concentration Games	experimental group (used computer game) having a greater mean score ($F = 53.61$, $p < 0.01$). There was also a significant difference in retention of spelling scores ($F = 17.57$, $p < 0.01$).
(Wu, Chen, & Huang, 2014)	To investigate whether a situated, digital learning playground could improve communicative skills and intrinsic motivation, in comparison with ordinary teaching and a board game group.	N = 96 (high school); FLF; English; Mixed; modified intrinsic motivation inventory; pre-and post-one-on-one computer-based communication assessment; informal interviews. GL, Comp, St: N/A (digital learning playground developed by researchers using Game Maker V.8.0 and based on the setting from Fresco, a commercial board game)	Participants in the experimental group (digital learning playground) achieved significantly better communication ability ($F = 3.27$, $p = 0.000$), and significantly lower tension ($p = .03$). There wasn't a significant difference between groups in interest, effort, or value.
(Young & Wang, 2014)	To explore the potential of using a Game Embedded CALL system with automatic speech recognition technologies to provide learners with individual opportunities for English pronunciation learning and vocabulary acquisition.	N = 52 (7-9); FLF English; Mixed; pre-and post-tests for speaking and vocabulary, perceived learning scales modified based on the Motivation-and-Attitudes-toward-Learning-English-Scale-for-Children (MALESC), semi-structured interviews, and a delayed vocabulary retention test. GL, Mobile, Ed: Game Embedded CALL system (GeCALL system; developed by researchers)	ANCOVA results show a significant difference in the pronunciation post-test scores with the experimental group (drill and game-based) performing better than the control group (drill practice) ($p = 0.03$), but the learners in the experimental group did slightly worse in the delayed vocabulary retention test. The game-based speaking activity also enhanced speaking motivation.

*Methods: number of participants (age); context; target language; data type; any additional information about the research design and measures.
FL = formal learning environment without game facilitation; FLF = formal learning environment with game facilitation; IL = informal learning environment; Quan = quantitative; Qual = qualitative.

**Game details: game purpose; platform used; game genre: name of game.
GL = game for learning; EG = entertainment game; Comp = Computer; Ed= educational; MG = educational mini-games; RPG = role-play; Ad = adventure; St = strategy; Sim = simulation.

Table A2
Quasi-experimental studies

Article	Aims	Methods* Game details**	Outcomes
(Hong, Han, Kim, Bae, Kim, & Renshaw, 2016)	To explore whether subjects' English language abilities after playing the HoDoo English game would be associated with increased brain functional connectivity in the areas of the brain involved in the language production and understanding networks.	N= 12 (9-10); unknown context, possibly IL; English; Mixed; pre-and post-tests using an fMRI scan, English ability and pragmatic skills assessment (qualitative?) GL, Comp, RPG: HoDoo English Game (online)	There was no significant difference in proficiency level ($p = .20$), but their total pragmatic skills improved $F(1,11) = 7.54$, $p = .02$). In addition, the twelve weeks of game play intensified the neural networking of Broca's area (language production) with the left middle frontal gyrus, as well as the neural networking of Wernick's area (understanding) with the left parahippocampal gyrus and the right medial frontal gyrus.

(Hong, Hwang, Tai, & Lin, 2017)	To analyze correlations between intrinsic motivation, self-efficacy, flow experience, and learning progress in recognizing Chinese radicals.	N = 73 (M = 10.62); FL; Chinese; Quan, single-group; questionnaire on motivation, self-efficacy, and flow, pre- and post-tests on Chinese radicals GL, Comp, Ed: N/A (researchers developed a Chinese radical learning game to be played with gesture-based interactions)	Intrinsic motivation, online learning self-efficacy, flow experience, and learning progress were all positively correlated. The mediating factors also predicted the degree of learning progress, and Chinese learning intrinsic motivation predicted online learning self-efficacy.
(Mazaji & Tabatabaei, 2016)	To investigate the effect of digital games on vocabulary acquisition of low-proficiency Iranian male and female EFL learners.	N = 60 (8-12); FL; English; Quan; Nelson language proficiency test, vocabulary pre- and post-tests based on students' course book. GL, Comp, MG: Polygot and speedy game	The digital game enhanced the vocabulary of the participants with the experimental group (game play) outperforming the control group (course books) (p = .000), and boys outperformed girls in the experimental group (p = .02).
(Pennala, Richardson, Ylinen, Lyytinen, & Martin, 2014)	To test the effectiveness of computer-assisted training (Graphogame) of Finnish phonemic length to improve accuracy.	N = 4 (7); FLF; Finnish; Quan; Wilcoxen signed-rank tests, pre- and post-tests GL, Comp, Ed: Graphogame	Observed differences between streams, words and pseudo-words indicate learning occurred and children applied the word representations they had learned to pseudo-words. Two of the four children seemed to benefit from the training at the end-assessment, as shown by the follow-up assessment.
(Hwang, G., Hsu, Lai, & Hsueh, 2017)	To examine anxiety, learning behaviors, motivation and performance in students' English learning through problem-based gaming.	N = 77 (9th grade, M = 15); FL; English; Quan; pre- and post-test of English listening skills based on the General English Proficiency Test, and questionnaires of learning motivation and English anxiety. GL, Comp, RPG: N/A (researchers developed a problem-based English listening game using RPG maker)	The experimental group (game play) performed significantly better than the control group (traditional teaching) in English listening ($F = 17.53$, $p < 0.001$, $n^2 = 0.19$). The same can be said for learning motivation ($F = 26.24$, $p < 0.001$, $n^2 = 0.26$). However, there was no significant difference in English-related anxiety.
(Sandberg, Maris, & Hoogendoorn, 2014)	To explore the added value of intelligent adaptation combined with game elements by testing two versions of a mobile game.	N = 106 (8-9.5); FLF; English; Quan; pre- and post-vocabulary tests based on SOPA and Peabody Picture Vocabulary Test, and passive vs. active word knowledge through an interview. GL, Mobile, Ad: Mobile English Learning (MEL)-original and MEL-enhanced (original game characteristics applied) developed by researchers	The students in the experimental condition (MEL-enhanced; passive: M = 20.73; active: M = 42.16) outperformed the children from the control group (MEL-original; passive: M = 28.14; active: M = 32.02), despite not spending significantly more time playing the game. Motivation to play the game may have been more extrinsic than intrinsic.
(Utku & Dolgunsöz, 2018)	To examine the effect of online vocabulary games on teaching new words to young learners of EFL and the learner perceptions.	N = 46 (11-13); FLF; English; Mixed for triangulation; pre- and post-recognition-production tests, and semi-structured interviews. GL, Comp, MG: matching game, word search, Canon Volley Sea Battle Game, crossword puzzle, spelling game, two board games – crocodile board game, pirate board game (eslgamesworld.com)	The results of the recognition ($F(1,44) = 4.485$, $p = .040$, $d = .41$.) and production ($F(1,44) = 7.620$, $p = .008$, $d = .97$) tests showed that experimental group (online games) significantly outperformed the control group (traditional teaching) in terms of vocabulary gains. The results of the semi-structured interviews also supported the quantitative results indicating that online vocabulary games increased students' motivation.
(Yang, Quadir, & Chen, 2016)	To investigate how the badge mechanism in DGBL can enhance users' self-efficacy in EFL.	N = 50 (9-10); FL; English; Quan; adapted self-efficacy questionnaire, and English learning performance pre- and post-tests based on the 3rd grade English curriculum.	The results showed that the badge mechanism had a significant positive influence on the learners' self-efficacy $t(49) = -2.303$, $p = .026$) and English learning performance ($t(49) = -2.728$, $p = 0.009$)

		GL, Comp, MG: (researchers developed a digital game with the badge mechanisms)	when comparing pre- and post-tests. Self-efficacy also affected English learning performance with students with higher self-efficacy performing better than those with lower self-efficacy

*Methods: number of participants (age); context; target language; data type; any additional information about the research design and measures.
FL = formal learning environment without game facilitation; FLF = formal learning environment with game facilitation; IL = informal learning environment; Quan = quantitative; Qual = qualitative.

**Game details: game purpose; platform used; game genre: name of game.
GL = game for learning; EG = entertainment game; Comp = Computer; Ed= educational; MG = educational mini-games; RPG = role-play; Ad = adventure; St = strategy; Sim = simulation.

Table A3
Other studies: Correlational, action research, design-based research, case study, and observational

Article	Aims	Methods*	Outcomes
		Game details**	
(Butler, Someya, & Fukuhara, 2014)	To examine the relationship between learners' use of online games and their performance in a related assessment, as well as the variance across age and gender.	N= 3,945 (4-12); IF; English; Quan, Correlational; descriptives of game use, and mock test and Jido-Eiken test scores to measure English performance.	Games children played with relative frequency shared a set of features: challenge, mystery, control, and multiple players. The frequency of plays decreased as the age increased in many games, but there weren't significant differences in gender. There were varying relational patterns between young learners' game scores and English performance depending on the games and difficulty levels of the assessment, meaning the more attractive games didn't necessarily yield strong learning results.
		GL, Comp, MG: Jido-Eiken learning game	
(Dalton & Devitt, 2016)	To assess the potential of using a 3D virtual environment (3DVE) for learning Irish by applying task-based language learning and cooperative learning principles.	N = 25 (9-10) for intervention; N = 15 (10-11) for later reflection component; FLF; Irish; Mixed (quantitative corpus analysis methods applied to the qualitative data to enhance Qual data), Action research; Pre- and post-motivational questionnaires (Attitude/Motivation Test Battery) and language tests, and interviews.	The motivation test showed children value the Irish language, but there were no significant gains from the language post-test, possibly due to the intervention being short. The most frequently occurring theme from the interviews was goal orientation, followed by construction, consumerism, travel, violence and social interaction. This reveals children may prefer a game environment with clear goals.
		EG, Comp, Sim: Open Sim	
(Dourda, Bratitsis, Griva, & Papadopoulou, 2014)	To test foreign language learning from an educational geography computer game.	N = 17 (11-12); FLF; English; Mixed, Case study; pre-intervention survey of digital habits and learning preferences, pre- and post-knowledge test on geography and English vocabulary, journal writing, game logs, observations, and a post-questionnaire on satisfaction/feedback.	All students improved on the knowledge test (M= 30% increase). Qualitative data confirmed that student vocabulary improved and the words were spelled correctly. Students seemed to comprehend texts more over time and their use of reading strategies was enhanced. The learning strategies used were also recorded: 65% used memory

		EG, Comp, Ad: Whodunit	strategies, 76% used cognitive strategies, 82% used social strategies, and 100% used compensation strategies. The students also reported satisfaction and enjoyment from the game and reported cooperation as one of the most positive aspects of the application.
(Lan, 2015)	To develop virtually immersive EFL learning contexts for EFL learners in Taiwan and to evaluate the effects on learners' EFL learning.	N = 132 (4-6 grade); FLF; English; Mixed, 2-iteration action research; iteration 1: field observation of behaviors and problems encountered, and teachers' comments; iteration 2: English performance test. EG, Comp, Sim: Second Life (researchers designed the virtual contexts)	Iteration 1: Almost all participants love learning in the English virtual contexts, most were excited when they earned presents. Iteration 2: Overall learning process was improved. Participants made significant improvement in sentences and conversation ($t=-3.242$, $p < 0.01$), but the improvement in vocabulary wasn't significant.
(Reitz, Sohny, & Lochmann, 2016)	Testing the ability to intrinsically evoke language communication training in a second language within a VR Game and whether this correlates with language requirements of a given curriculum in order to improve language acquisition and training.	N = 26 (12-34); IL; English; Mixed, N/A (action research/ design-based?); Pre-questionnaire on computer use behaviour and self-assessment of English, a motivation questionnaire (QCM) mid-game (7 point Likert), transcription of mid-game conversations. GL, Comp, N/A Sim?: Haunted (developed as joint project between teachers and game designers)	Players felt challenged (M= 5.02, SD= 1.14), they were interested in the game (M=5.1, SD = 1.03), and there was relatively low reported anxiety (M = 3.3, SD = 1.53), which reveals the game was motivating. In addition, most players produced correct grammatical structures (asking questions, using can and can't, and prepositions of movement), but most mistakes were not corrected by other players. The game also gave players confidence to communicate.
(Sundqvist & Wikstrom, 2015)	To investigate the relation between out-of-school digital gameplay and in-school L2 English vocabulary measures and grading outcomes.	N = 80 (15-16); IL; English; Mixed, correlational; questionnaire, language diaries, vocabulary tests (productive and vocabulary levels tests), assessed essays, and grades using an observational post-hoc design. EG, N/A, N/A: variety of games	Three digital game groups were created based on frequency of gameplay. Those who played most frequently (DGG3) had the highest rated essays, most advanced vocabulary in essays, and highest grades, followed by non-gamers (DGG1) and then moderate gamers (DGG2). DGG1 performed significantly better on the vocabulary test ($p = 0.000$), followed by DGG2 and DGG1 where there was no significant difference.
(Shahriarpour & kafi, 2014)	To examine the role of digital games on vocabulary learning and language learning motivation.	N = 25 (14-16); IL; English; Qual, observational; interviews and observations. EG, N/A, Ed: L.A.Noire digital game	Using the game to practice vocabulary enhanced students' ability to acquire words and enhanced motivation.
(Wong & Hsu, 2016)	To design novel learning activities to foster orthographic awareness, which includes understanding the ways in which components can be combined to form Chinese characters correctly, and also the commonly used structures in these formations	N = 31 (9); FLF; Chinese; Mixed for triangulation, design-based research; pre- and post-tests on character formation, observations, video and audio recordings, and post-interviews. GL, Mobile, MG: Chinese-PP	The flexible grouping approach stimulated peer interaction after several rounds of games. Post-test scores were significantly better than the pre-test ($t = -4.38$; $p < .05$), and there was a significant difference between learning effectiveness of sensing-style students than intuitive-style students ($t = 2.70$; $p < .05$). There was no significant difference between students of active and reflective learning styles though ($t = -.06$; $p > 0.05$).

*Methods: number of participants (age); context; target language; data type; any additional information about the research design and measures.

FL = formal learning environment without game facilitation; FLF = formal learning environment with game facilitation; IL = informal learning environment; Quan = quantitative; Qual = qualitative.

**Game details: game purpose; platform used; game genre: name of game.
GL = game for learning; EG = entertainment game; Comp = Computer; Ed= educational; MG = educational mini-games; RPG = role-play; Ad = adventure; St = strategy; Sim = simulation.

References

All, A. (2014). Measuring Effectiveness in Digital Game-Based Learning: A Methodological Review. *International Journal of Serious Games, 1*(2) doi:http://dx.doi.org/10.17083/ijsg.v2i4.98

Beavis, C., Rowan, L., Dezuanni, M., McGillivray, C., O'Mara, J., Prestridge, S., Stiler-Hunt, C., Thompson, R., & Zagami, J. (2014). Teachers' Beliefs about the Possibilities and Limitations of Digital Games in Classrooms. *E-Learning and Digital Media, 11*(6), 569-581. doi:http://10.2304/elea.2014.11.6.569

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, M., Ninaus, M., Ribeiro, C., & Pereira, J. (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:https://doi.org/10.1016/j.compedu.2015.11.003

Chiu, Y., Kao, C., & Reynolds, B. L. (2012). The relative effectiveness of digital game-based learning types in English as a foreign language setting: A meta-analysis. *British Journal of Educational Technology, 43*(4), E107. doi:https://doi.org/10.1111/j.1467-8535.2012.01295.x

Godwin-Jones, R. (2018). Chasing the butterfly effect: Informal language learning online as a complex system. *Language Learning & Technology, 22*(2), 8-27. doi:https://doi.org/10125/44643

Hainey, T., Connolly, T. M., Boyle, E. A., Wilson, A., & Razak, A. (2016). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 102*, 202-223. doi:https://doi.org/10.1016/j.compedu.2012.03.004

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning.*Computers in Human Behavior, 54*, 170-179. doi:https://doi.org/10.1016/j.chb.2015.07.045

Herz, J. C. (1997). *Joystick nation: How videogames gobbled our money, won our hearts, and rewired our minds* (1st uppl.). Boston: Little, Brown and Company.

Hung, H., Yang, J. C., Hwang, G., Chu, H., & Wang, C. (2018). A scoping review of research on digital game-based language learning. *Computers & Education, 126*, 89-104.

doi:https://doi.org/10.1016/j.compedu.2018.07.001

Hwang, G. (2012). Advancements and trends in digital game-based learning research: a review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology*, *43*(1), E10. doi:https://doi.org/10.1111/j.1467-8535.2011.01242.x

Klopfer, E., Osterweil, S., & Salen, K. (2009). *Moving learning games forward.* (). Cambridge, MA: Education Arcade.

Kurt D. Squire. (2008). Video Games and Education: Designing Learning Systems for an Interactive Age. *Educational Technology*, *48*(2), 17-26. Hämtad från https://www.jstor.org/stable/44429558

Li, M., & Tsai, C. (2013). Game-Based Learning in Science Education: A Review of Relevant Research. *Journal of Science Education and Technology*, *22*(6), 877-898. doi:https://doi.org/10.1007/s10956-013-9436-x

Moher, D., Liberati, A., Tetzlaff, J., & G Altman, D. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, *6*(7), e1000097. doi:https://doi.org/10.1371/journal.pmed.1000097

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement.*Systematic Reviews*, *4*(1), 1. doi:http://10.1186/2046-4053-4-1

Oblinger, D. (2004). The Next Generation of Educational Engagement. *Journal of Interactive Media in Education*, *1*, Article 10. doi:https://doi.org/10.5334/2004-8-oblinger

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:https://doi.org/10.1016/j.compedu.2009.04.001

Perttula, A. (2017). Flow experience in game based learning – a systematic literature review. *International Journal of Serious Games*, *4*(1) doi:http://dx.doi.org/10.17083/ijsg.v4i1.151

Peterson, M. (2010). Massively multiplayer online role-playing games as arenas for second language learning. *Computer Assisted Language Learning*, *23*(5), 429-439.

doi:https://doi.org/10.1080/09588221.2010.520673

Peterson, M. (2016). The use of massively multiplayer online role-playing games in CALL: an analysis of research. *Computer Assisted Language Learning, 29*(7), 1181. doi:https://doi.org/10.1080/09588221.2016.1197949

Prensky, M. (2001). The Digital Game-Based Learning Revolution. *Digital game-based learning* (). United States: McGraw-Hill Publishing.

Shaffer, D. W., Squire, K. R., Halverson, R., & Gee, J. P. (2005). Video Games and the Future of Learning. *Phi Delta Kappan, 87*(2), 105-111. doi:https://doi.org/10.1177/003172170508700205

Squire, K. D. (2002). Cultural Framing of Computer/Video Games. *Game Studies: The International Journal of Computer Game Research, 2*(1) Hämtad från http://www.gamestudies. org/O 702/squire/

Wastiau, P., Kearney, C., & Van den Berghe, W. (2009). *How are digital games used in school?* (). Brussels: European Schoolnet.

Wilkinson, P. (2016). A brief history of serious games. I Dörner, R., Göbel, S., Kickmeier-Rust, M., Masuch, M., Zweig, K.A., R. Dörner, S. Göbel, M. Kickmeier-Rust, M. Masuch & K. A. Zweigh (Red.), *Entertainment computing and serious games* (1 uppl., ss. 17-41). Cham: Springer. doi:10.1007/978-3-319-46152-6

Yip, F. W., & Kwan, A. C. M. (2006). Online vocabulary games as a tool for teaching and learning English vocabulary. *Educational Media International, 43*(3), 233-249. doi:https://doi.org/10.1080/09523980600641445

Young, M., F, Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & Yukhymenko, M. (2012). Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. *Review of Educational Research, 82*(1), 61-89. doi:https://doi.org/10.3102/0034654312436980

Zuengler, J., & Miller, E. R. (2006). Cognitive and Sociocultural Perspectives: Two Parallel SLA Worlds? *TESOL Quarterly, 40*(1), 35-58. doi:http://10.2307/40264510

Coded Papers

Butler, Y. G., Someya, Y., & Fukuhara, E. (2014). Online games for young learners' foreign language learning. *English Language Teaching Journal, 68*(3), 275}. doi:https://doi.org/10.1093/elt/ccu008

Cornillie, F., Van Den Noortgate, W., Van den Branden, K., & Desmet, P. (2018). Examining Focused L2 Practice: From in Vitro to in Vivo. *Language Learning & Technology, 23*(3), 10-26. Hämtad från https://eric.ed.gov/?id=EJ1142262

Dalton, G., & Devitt, A. (2016). Action research Irish in a 3D world: Engaging primary school children. *Language Learning & Technology, 20*(1), 21-33. Hämtad från http://llt.msu.edu/issues/february2016/action.pdf

Dourda, K., Bratitsis, T., Griva, E., & Papadopoulou, P. (2014). Content and language integrated learning through an online game in primary school: A case study *Electronic Journal of E-Learning, 12*(3), 243-258. Hämtad från http://www.ejel.org/volume12/issue3/p243

Ebrahimzadeh, M. (2017). Readers, Players, and Watchers: EFL Students' Vocabulary Acquisition through Digital Video Games. *English Language Teaching, 10*(2), 1-18. doi:http://dx.doi.org/10.5539/elt.v10n2p1

Ebrahimzadeh, M., & Alavi, S. (2016). Motivating EFL students: E-learning enjoyment as a predictor of vocabulary learning through digital video games. *Cogent Education, 3*(1), 1255400. doi:http://dx.doi.org/10.1080/2331186X.2016.1255400

Hong, J. S., Han, D. H., Kim, Y. I., Bae, S. J., Kim, S. M., & Renshaw, P. (2016). English language education on-line game and brain connectivity. *ReCALL, 29*(1), 3-21. doi:http://doi.org/10.1017/S0958344016000173

Hong, J., Hwang, M., Tai, K., & Lin, P. (2017). Intrinsic motivation of Chinese learning in predicting online learning self-efficacy and flow experience relevant to students' learning progress. *Computer Assisted Language Learning, 30*(6), 552-574. doi:http://doi.org/10.1080/09588221.2017.1329215

Hsu, T. (2017). Learning English with Augmented Reality: Do learning styles matter? *Computers & Education, 106*, 137-149. doi:http://doi.org/10.1016/j.compedu.2016.12.007

Hwang, G., Hsu, T., Lai, C., & Hsueh, C. (2017). Interaction of problem-based gaming and learning anxiety in language students' English listening performance and progressive behavioral patterns. *Computers & Education, 106*, 26-42. doi:http://dx.doi.org/10.1016/j.compedu.2016.11.010

Hwang, G., & Wang, S. (2016). *Single loop or double loop learning: English vocabulary learning performance and behavior of students in situated computer games with different guiding strategies*doi:http://doi.org/10.1016/j.compedu.2016.07.005

Hwang, W., Shih, T. K., Ma, Z., Shadiev, R., & Chen, S. (2016). Evaluating listening and speaking skills in a mobile game-based learning environment with situational contexts. *Computer Assisted Language Learning, 29*(4), 639-657. doi:https://doi.org/10.1080/09588221.2015.1016438

Lan, Y. J. (2015). Contextual EFL learning in a 3D virtual environment. *Language Learning & Technology, 19*(2), 16-31. Hämtad från http://llt.msu.edu/issues/june2015/action.pdf

Mazaji, B., & Tabatabaei, O. (2016). Effects of digital games on vocabulary acquisition of Iranian low proficiency male and female EFL learners *Modern Journal of Language Teaching Methods, 6*(4), 151-163.

Pennala, R., Richardson, U., Ylinen, S., Lyytinen, H., & Martin, M. (2014). Computer game as a tool for training the identification of phonemic length. *Logopedics Phoniatrics Vocology, 39*(4), 149-158. doi:http://doi.org/10.3109/14015439.2013.810302

Reitz, L., Sohny, A., & Lochmann, G. (2016). VR-based gamification of communication training and rral examination in a second language. *International Journal of Game-Based Learning, 6*(2), 391-404. doi:http://dx.doi.org/10.1080/14681811.2015.1123147

Salehi, H. (2018). Effects of Using Instructional Video Games on Teaching English Vocabulary to Iranian Pre-Intermediate EFL Learners. *International Journal of Learning and Change, 31*, 72-89. doi:http://dx.doi.org/10.1080/09588221.2017.1382536

Sandberg, J., Maris, M., & Hoogendoorn, P. (2014). The added value of a gaming context and intelligent adaptation for a mobile learning application for vocabulary learning. *Computers & Education, 76*, 119-130. doi:http://doi.org/10.1016/j.compedu.2014.03.006

Shahriarpour, N., & kafi, Z. (2014). On the effect of playing digital games on Iranian intermediate EFL learners' motivation toward learning English vocabularies. *Procedia - Social and Behavioral Sciences, 98*, 1738-1743. doi:http://doi.org/10.1016/j.sbspro.2014.03.601

Shokri, H., & Abdolmanafi-Rokni, S. J. (2014). The Effect of Using Educational Computer Games on Recall and Retention of Spelling in Iranian EFL Learners. *International Journal of Applied Linguistics & English Literature, 3*(6), 169-175. doi:http://doi.org/10.7575/aiac.ijalel.v.3n.6p.169

Sundqvist, P., & Wikstrom, P. (2015). Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *System, 51*, 65-76. doi:http://doi.org/10.1016/j.system.2015.04.001

Utku, Ö, & Dolgunsöz, E. (2018). Teaching EFL vocabulary to young digital natives through Online. *International Online Journal of Education and Teaching, 5*(1), 115-130. Hämtad från http://iojet.org/index.php/IOJET/article/view/228

Wong, L., & Hsu, C. (2016). Effects of learning styles on learners' collaborative patterns in a mobile-assisted, Chinese character-forming game based on a flexible grouping approach. *Technology, Pedagogy and Education, 25*(1), 61-77. doi:http://doi.org/10.1080/1475939X.2014.963661

Wu, C., Chen, G., & Huang, C. (2014). Using Digital Board Games for Genuine Communication in EFL Classrooms. *Educational Technology Research and Development, 62*, 209-226. doi:http://doi.org/10.1007/s11423-013-9329-y

Yang, J. C., Quadir, B., & Chen, N. (2016). Effects of the Badge Mechanism on Self-Efficacy and Learning Performance in a Game-Based English Learning Environment. *Journal of Educational Computing Research, 54*(3), 371-394. doi:http://doi.org/10.1177/0735633115620433

Young, S. S., & Wang, Y. (2014). The Game Embedded CALL System to Facilitate English Vocabulary Acquisition and Pronunciation. *Journal of Educational Technology & Society, 17*(3), 239-251. Hämtad från http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=1&sid=15fb597c-552a-4c20-9cac-c3a1b168c85e%40pdc-v-sessmgr03