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

Emmanuel O. Acquah*, Heidi T. Katz

Faculty of Education and Welfare Studies, Åbo Akademi University, Vaasa, Finland

ARTICLE INFO

ABSTRACT

The aim of this systematic literature review was to examine the empirical evidence for the effectiveness of digital games on second language learning between 2014 and 2018, with a focus on participants 6–18 years old. The initial search yielded 578 results, from which a total of 26 articles were included in the final content analysis. The analysis of the included studies revealed: (1) the majority of studies were conducted with a mixed methods design; (2) most studies used computers as the gaming platform; (3) the most common game genre was educational games or educational mini games; (5) most games were designed for learning; (4) research was mainly conducted in East Asia and the Middle East; (5) the primary context of study was within a formal learning environment; and (6) the target language was usually English. Further analysis suggests that digital learning games (DLGs) may benefit players' language acquisition, affective/psychological state, contemporary competences, and participatory behavior. An inductive analysis revealed six key game features highlighted within the studies that influenced the outcomes: ease-of-use, challenge (at one's zone of proximal development), rewards and feedback, control or autonomy, goal-orientation, and interactivity. In addition to game features, associations between context and outcomes were also explored: studies conducted within a formal learning environment, with or without teacher facilitation, resulted in mostly positive language acquisition results, meaning DLGs can be implemented successfully within schools. Based on the overall findings, it is clear that DLGs are an effective tool, but future research should explore how they can best be implemented in the classroom setting.

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, 2008), but the existing literature related to digital game-based language learning (DGBLL) is still scarce (Hung, Yang, Hwang, Chu, & Wang, 2018; Peterson, 2016). This systematic review expands the existing literature by framing discussions around how DGBLL can be implemented in school, and what designers need to consider when creating digital learning games (DLGs) for primary through high-school age students.

* Corresponding author. Faculty of Education and Welfare Studies, PO Box 311, 65101, Åbo Akademi University, Vaasa. Finland.
 E-mail address: emmanuel.acquah@abo.fi (E.O. Acquah).

1.1. 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, iPads, 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 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 (Hamari et al., 2016; Klopfer, Osterweil, & Salen, 2009; Oblinger, 2004; Prensky, 2001; Shaffer, Squire, Halverson, & Gee, 2005; 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 often 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). This is especially true for L2 learning, which has been most often used, and associated with, positive outcomes compared to more mixed results in other subject areas (Wastiau, Kearney, & Van den Berghe, 2009; Young et al., 2012). Research attributes the success of DLGs for L2 acquisition to an increased exposure to the target language and enhanced engagement (Aghlara & Tamjid, 2011; Ashraf, Motlagh, & Salami, 2014). In addition, language learning is more effective when situated and when learners can be social, two components found in DGBL (Gee, 2004; Peterson, 2016).

1.2. Age effect on learning languages

According to the much debated critical period hypothesis (CPH) in L2 acquisition, there is an ideal window in which optimal L2 acquisition occurs: the younger the better (Sang, 2017). A recent large-scale study provided the first direct estimate of the critical period (CP), finding learning ability is preserved much later than originally thought (until 17–18 years old), but to reach native-level proficiency, learners should start between 10 and 12 years old (Hartshorne, Tenenbaum, & Pinker, 2018). Other researchers have discussed the age effect on learning languages, addressing the CP as well as highlighting differences in L2 learning behavior between children and adults (Abdullah & Akhter, 2015; Sang, 2017). Therefore, it is essential to introduce L2 instruction early, and for primary through secondary educators to have a language curriculum that best meets the needs of NetGen'ers. Considering the CP and in order to assess whether DLGs can be used to enhance L2 learning in primary and secondary schools, this review focused on children and adolescents during the CP, as opposed to university students or adults.

Previous DLG reviews have also set similar age limitations: Young et al. (2012) examined the use of video games as part of the K-12 curricula; Boyle et al. (2016) and Connolly, Boyle, MacArthur, Hainey, and Boyle (2012) looked at participants over the age of 14; and Hainey, Connolly, Boyle, Wilson, & Razak, 2016 focused on primary education in order to address the gap left by Boyle et al. (2016) and Connolly et al. (2012). By focusing on digital games for child and adolescent L2 learning, the research findings can more confidently be applied to instruction and curriculum in primary and secondary schools. Unfortunately, teachers often struggle to integrate DLGs into their classroom instruction due to time constraints, struggling to connect the games to the curriculum, and attitudes of the school and parents (Wastiau et al., 2009). In addition, the use of DLGs for L2 instruction must be very intentional for optimal learning to occur; high interactivity, improper facilitation by teachers or institutions, or poor game choice can detract from the learning (deHaan, 2011; Klimova & Kacet, 2017). To overcome these challenges researchers must provide explicit and applicable age-appropriate knowledge; game designers should consider the research on learning and gaming to develop games specifically for classroom use; and progressive educators need to support teachers in becoming familiar with DLGs.

2. The present study

This review was conducted in consideration of the positive research on DLGs and systematic reviews on the topic. Previous reviews on DLGs have looked at positive affect in specific subject areas (Chiu, Kao, & Reynolds, 2012; Hung et al., 2018; Kim, 2018; Klimova & Kacet, 2017; Li & Tsai, 2013; Papastergiou, 2009; Peterson, 2016; Young et al., 2012; Yudintseva, 2015), the experience of flow or engagement (Perttula, 2017), trends in research (Hwang, 2012), and research methods used (All, 2014). While the aforementioned reviews were considered throughout this review process, two reviews were more intensely studied and they provided a framework to conduct this review. First, Connolly et al. (2012) examined 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 et al., 2016). While they did not specifically look at language, the current review was influenced by their search terms and two coding categories: game genre (Herz (1997) classification) and intent of game (Boyle et al., 2016; Connolly et al., 2012; Hainey et al., 2016). The updated review by Boyle et al. (2016) will be referred to most in this study.

For coding outcomes and as an overall starting point, this review was influenced by Hung et al. (2018) who published a scoping overview of the empirical evidence on the use and impacts of digital games in language learning from 2007 to 2016. Hung et al. (2018) searched 10 journals for research on DGBLL at all ages, including L1 learning. They argued that there is a need for more

focused literature in order to understand how DLGs can be utilized in specific content areas (Hung et al., 2018). In answering that call, this review narrows the search and focus by looking at primary through secondary school age participants; excluding L1 learners; analyzing the context of gameplay; using different, more specific search terms; excluding games played on video consoles, as they aren't typically used in classrooms; and observing common game features that are discussed across articles. This review broadens their search only by including more databases (Hung et al. (2018) used the Web of Science database), and by overlapping in the years reviewed (they reviewed articles from 2007 to 2016). The findings can be used to inform educators (of primary through high school students), researchers, and game designers about outcomes from DGBLL, key game features, and the types of games already available to use for language learning.

2.1. Defining terms

Digital games used for the purpose of learning have been discussed using a variety of terms: digital education game (DEG), serious game, DGBL, DGBLL, game for learning, and 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 Boyle et al. (2016), Connolly et al. (2012) and Hainey et al. (2016) in their systematic reviews when discussing the game type based on developer intention: *game for learning* (GL) and *entertainment game* (EG). They defined EGs as "pre-made, COTS [commercial off-the-shelf] 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 all their studies, serious games fell under the category of GL because the categories were too close to be distinct (Boyle et al., 2016; Connolly et al., 2012; 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 comprises 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 et al. (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).

In addition to defining terms that will be used to discuss DLGs, it is also important to clarify how *outcomes* are understood in this review. At a basic level, DLGs function as tools for learners, but the purpose for using DLGs rather than alternative teaching methods is that they result in other types of positive outcomes and are often better matched to how children learn today (Roschelle, Pea, Hoadley, Gordin, & Means, 2000). According to cognitive research, learning is most effective when it involves active engagement, group participation, frequent interaction and feedback, and connections to real-world contexts (Roschelle et al., 2000). Therefore, it isn't enough to say that DLGs impact language learning outcomes, but it is necessary to consider the kind of experience they provide the learner because the learning outcomes depend on a variety of factors (Roschelle et al., 2000). It also is important to consider the purpose and value of education. Modern education should enable students to grow, problem solve, think critically, and better understand one's role in society (Noddings, 2015). In considering effective learning and the purpose of education, *outcomes* in this article refer to anything that can enable a person to grow academically, socially, psychologically, and toward accomplishing their own goals (Noddings, 2015).

2.2. Aim and research questions

This review aims to provide a better understanding of how digital games can be used to enhance language learning for students, especially within classroom settings given that teachers desire to learn more about how to use DLGs as tools (Wastiau et al., 2009). This review is an attempt to address that need by giving 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 are searched. In addition, the choice to focus on studies with participants from age 6 to 18 is threefold: (1) a primary focus in this study is on how teachers can use digital games as tools in primary through secondary language classrooms; (2) there is an age effect on learning languages (Abdullah & Akhter, 2015; Hartshorne et al., 2018; Sang, 2017), meaning the outcomes from using digital games, whether in a formal education setting or not, will likely be different after the CP (age 18); and (3) game designers need to consider the target group (age and setting), especially if the games are designed to be used specifically for language learning within a classroom of NetGen'ers. Given the aforementioned aims of this review, a search was conducted to answer the following research questions:

1. What empirical evidence is there concerning the positive impacts and outcomes of DGBLL on primary through high school-age children?
2. How do game characteristics and context influence DGBLL outcomes?

3. Method

3.1. 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.

3.2. 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 to 2018.
6. Written in English
7. Studies on children from primary school through high school (age 6–18)
8. Discusses L2 (second-language or foreign language) learning from playing a DLG
9. Discusses outcomes from gameplay

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

1. Board games, simulations (unless simulation is the game genre), and digital tools
2. Game platform: video console
3. Systematic reviews, meta-analyses, book chapters, position papers, and conference proceedings
4. L1 learning
5. Related to learning from the participant's designing games

Originally, mobile devices were under the exclusion criteria as they are not 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 games available on mobile devices can also be played on tablets. Studies using language applications that were not in a gaming format were still excluded, but it was decided that mobile devices could be included. There were also two studies where the game console was unknown, but these were included with reason: Shahriarpour and kafi (2014) used L.A. Noire, a game that could be played on a video console, but also on a computer; Sundqvist and Wikstrom (2015) researched informal out-of-school gameplay so it was likely that at least some gameplay occurred on a computer, tablet, or mobile phone. In addition, some studies included participants whose age range exceeded the 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.

3.3. 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.

3.4. Search terms

The search terms used were related to DLGs, outcomes, and language acquisition. The search included similar terms to those used by Boyle et al. (2016), 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: the terms specify only the use of digital games, and the terms are broader in order to expand the possible types of outcomes from DGBLL. While there is a clear coding scheme for outcomes presented in the results, the terms used in the search encompass a variety of outcomes, which supports how outcomes are understood in this study. 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 general 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.

3.5. 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 one or more criteria was absent or unclear in the title or abstract, the articles were included 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 sections, allowing articles to be added through snowballing.

In the second screening phase, the researchers met again to discuss how they would assess the aim, methods, results, and conclusion of each paper to determine inclusion. The researchers tested for inter-rater reliability with ten papers, finding they were in agreement 90% of the time based on Cohen (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 et al., 2015; Moher, Liberati, Tetzlaff, & G Altman, 2009). While this review couldn't be registered with Prospero (ineligible due to it not being related to health), per their recommendation, this review followed the PRISMA statement guidelines (see Moher et al., 2015), and as stated previously, the protocol was uploaded to OSF instead.

3.6. Data analysis

The two researchers conferred on how to code the papers, and then split the papers to code independently through content analysis. Problems in coding were discussed throughout the process and the researchers shared and reviewed articles when necessary.

The analyzed data included both manifest and latent content; manifest content is easily observed, while latent content is more difficult to analyze since the analyst must look below the surface at the underlying meaning (Bengtsson, 2016; Potter & Levine-Donnerstein, 1999). Target language, whether the game was developed by researchers, and country of research are examples of manifest content, since all authors directly stated the information, thus coding for these was straightforward. In addition, game genre, research methods, game type, context of gameplay, and game platform were at times manifest content, but in some articles these were not directly stated and had to be inferred based off of content cues and coder schema. In these cases, the coders used agreed-upon coding rules and discussed any questionable items. For example, Reitz, Sohny, and Lochmann (2016) never directly state their research design, but their study was coded as design/action based research due to the coder's interpretation of text such as, "The authors' central contribution is to answer [how the researchers can intrinsically evoke language communication training in a second language within VR] by presenting a game design template ..." (p.47).

While outcomes and game features were also sometimes stated directly, these were generally more open for interpretation especially in qualitative articles where findings weren't always explicit. Game outcomes did have a predetermined coding scheme though, and along with game genre, were coded deductively with some alterations being made after the coding began. Game features on the other hand were coded inductively, since the data had to be observed first before developing empirical generalizations (Bengtsson, 2016; Potter & Levine-Donnerstein, 1999). All other items were coded deductively without alterations. More specific details on how each category was coded are presented along with the results for those categories.
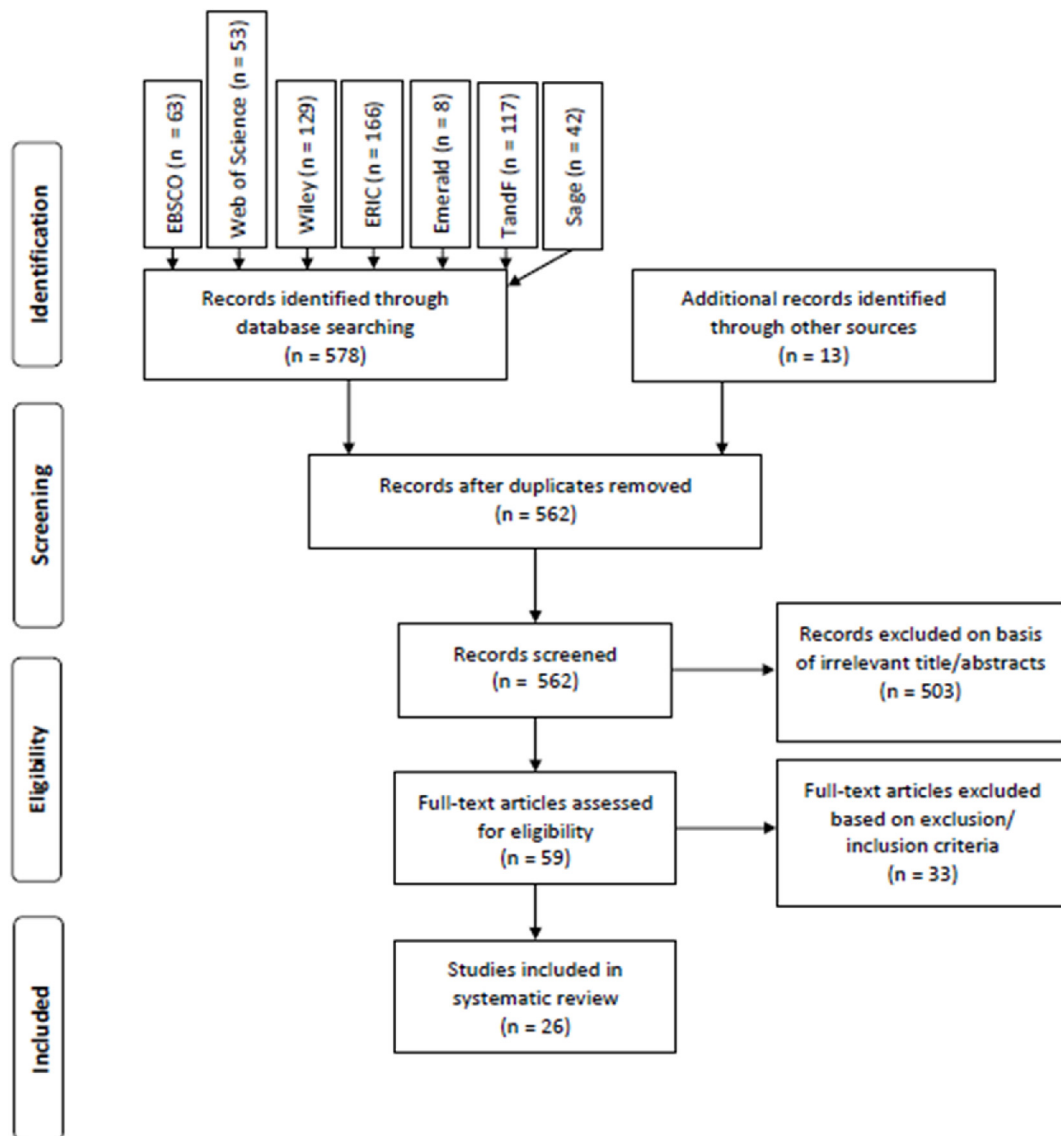
Fig. 1. Flow diagram of the screening and selection procedure.

4. Results

4.1. Overview of included papers

As seen in Figs. 1 and 26 out of the initial 578 identified papers 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 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%).

Fig. 2 shows the number of studies using each design, based on the number of participants in the study. The largest study was correlational, with 3945 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).

4.1.1. Research methods

The majority of studies used mixed methods (n = 14; 54%) for data collection, followed by quantitative (n = 11; 42%), and one
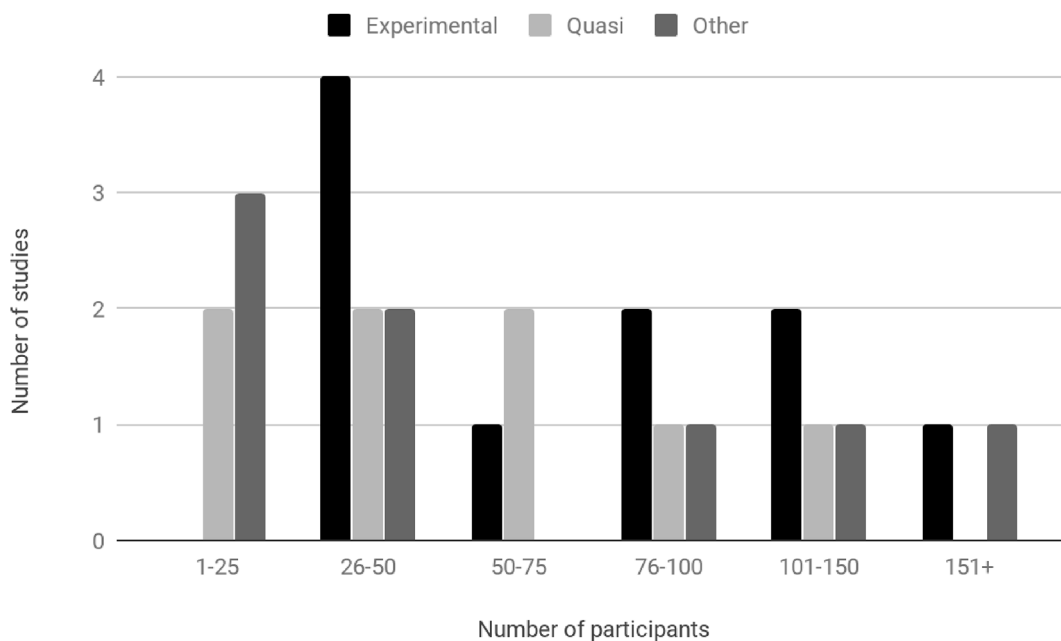
Fig. 2. Number of participants by study design.

qualitative study (4%). This finding is similar to that of Hung et al. (2018) who also reported the majority of studies being mixed methods (n = 25; 50%), followed by quantitative (n = 16; 32%) and then qualitative (n = 9; 18%). Given that the focus of this field of research is on both lived experiences and set outcomes, these are logical findings. Not only do researchers want to statistically report on learning outcomes, but they also want to understand how users experience the games. The mixed methods studies can therefore paint a clear picture on whether games lead to positive outcomes, and whether the players enjoyed the gameplay.

In general, the actual coding of the papers based on methods, as well as research design, posed a challenge. As mentioned in the analysis section, the methods or research design were often not 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 design. For instance, Hwang & Wang (2016) classified their study as experimental despite the fact that they didn't randomly select participants, but instead assigned two classes in a school into the two different groups. As a rule though, the classification given in the article was used, if available.

4.1.2. 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 (for each n = 1; 4%). From their included L2 studies, Hung et al. (2018) also found English was the most common target language (77%), with other studies focusing on German, Spanish, Chinese, Italian, and Japanese. Interestingly, Hung et al. (2018) classified Dalton and Devitt (2016) study on Irish learning in Ireland as L1, but the study was included in this review since within the context of the study (and typically in the Irish school context), Irish is a second language.

4.1.3. Game platform

Concerning game platforms, 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%). Computers were found to be the most common platforms by both Connolly et al. (2012) and Hung et al. (2018), though Boyle et al. (2016) reported videogames were the majority, followed by computers. Given that computers have been around the longest and that they function as writing tools, this fact isn't surprising especially since the majority of studies in this review took place within a formal learning environment. However, it was surprising that tablets and mobile phones weren't used more frequently as their portability has made them more popular (Newhouse, 2014). The portability of these platforms was even referred to in two of the three articles that used them (Hwang, Shih, Ma, Shadiev, & Chen, 2016; Sandberg, Maris, & Hoogendoorn, 2014). However, other than the fact that computers have been used in classrooms longer, it is also possible that articles using these platforms more often refer to games as "applications," a word that was purposefully excluded from the search.

4.1.4. 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 (27%; excluding Turkey), 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.
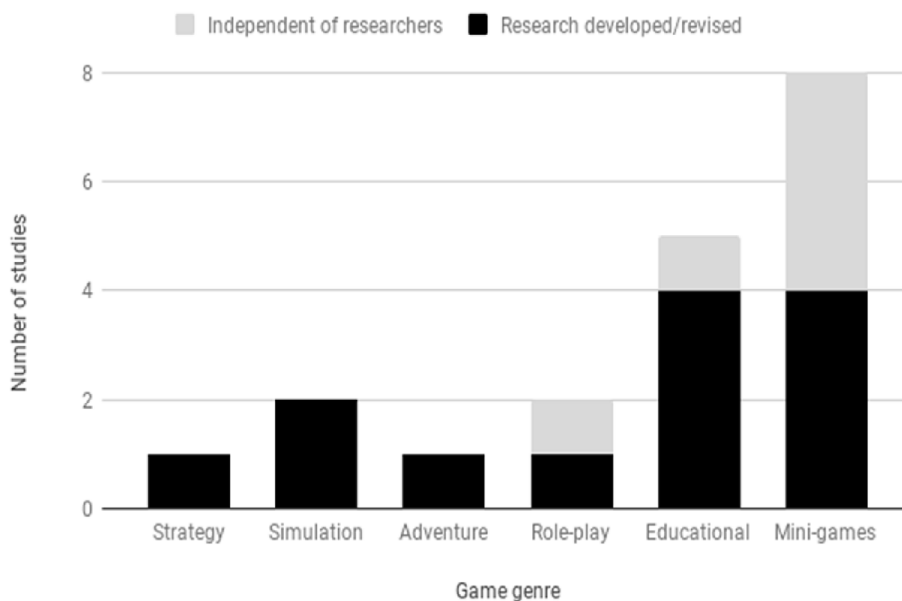
Fig. 3. Games for learning: Number of studies using each game genre and whether researchers developed or revised the games.

4.2. Game genre

As in the review by Boyle et al. (2016), Herz (1997) system was used in order to help code latent game genres deductively. The game genre was often left unclassified in the articles, and Herz (1997) descriptions acted as a guide for interpreting the text and then coding the genre. After the first attempt to code the articles by Herz (1997) system, it was decided that educational mini-games and educational games would be added, and thus inductive reasoning was used as well. This was due to the fact that the studies using educational mini-games often included a variety of games for learning, each a different game genre, 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 latent content 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).

The quote describes how to play, but doesn't provide a visual of what is actually on the screen. Descriptions like this don't fit into Herz (1997) system, which is why the new categories had to be created.

The majority of games used in these studies were classified as either educational mini-games (n = 8; 31%) or educational games (23%), while the fewest number of studies used games classified as strategy or adventure (for each n = 3; 12%). Fig. 3 and Fig. 4 present the number of games per genre and whether the authors of the study were involved in either developing the game or reworking a pre-existing GL (commonly seen with simulation games). Fig. 3 shows only the GLs and Fig. 4 displays the EGs. Sundqvist and Wikstrom (2015) studied out-of-school gameplay, which means the game genre and intention of the game (GL or EG) varied per person making them unknown, thus the study was excluded from Figs. 3 and 4. As is made clear in the figures, the majority of games were research developed (n = 15; 58%) and GLs (n = 19; 73%), with only 6 (23%) identified as EGs. Hung et al. (2018) had a similar finding, reporting only 36% of included studies utilizing COTS (EGs). On the contrary, Boyle et al. (2016) found 53% of the games were EGs, but this could be due to the fact that their focus wasn't solely on language learning.

None of the games used in the research articles were action, puzzle, fighting or sports, which is logical since these games may be more distracting, too violent, or not applicable as language learning tools. There were no clear patterns between game genre and outcomes, and not enough studies using each genre to draw conclusions. However, more specific game features highlighted in the studies were observed to explore impact on outcomes.

4.3. Game features

Game features were open-coded inductively based on the article's direct reference to a specific factor built into the game that impacted the player. One study found the badge mechanism, which provides instant feedback, positively impacted players' self-efficacy and learning performance (Yang, Quadir, & Chen, 2016). Butler, Someya, and Fukuhara (2014) noted four features shared by the mini-games children chose to play most frequently: challenge, mystery, control, and having multiple players. Autonomy (player control) - was supported in another study that found a positive correlation with vocabulary learning, while challenge had a negative
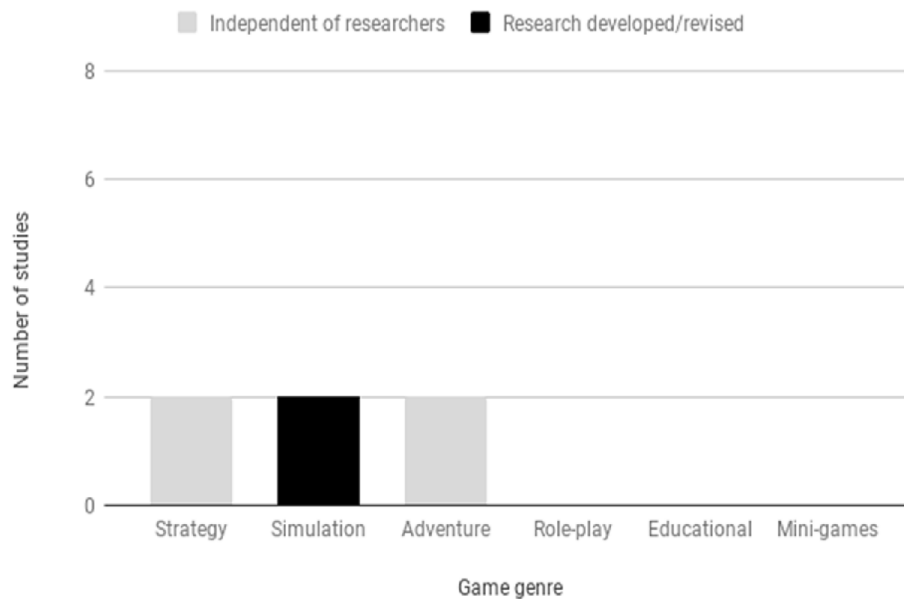
Fig. 4. Entertainment games: Number of studies using each game genre and whether researchers developed or revised the games.

correlation with vocabulary learning as it can sometimes cause anxiety (Ebrahimzadeh & Alavi, 2016). Several authors explain how games should provide an adequate amount of challenge, but not too much where the players cannot be successful (Butler et al., 2014; Ebrahimzadeh & Alavi, 2016; Hong, Hwang, Tai & Lin, 2017; Hsu, 2016; Reitz et al., 2016; Sandberg et al., 2014). This is a direct representation of Vygtosky's Zone of Proximal Development: the idea that effective learning occurs when a student is given the proper support to succeed, but is still challenged within his or her skill range (Vygotsky, 1978). The instant feedback, referred to previously in the study from Yang et al. (2016), adds the layer of support required to work in one's ZPD. This balance of skill and challenge can motivate players and place one in what Csikszentmihalyi (1990) refers to as a flow state: an optimal experience of high concentration where one finds great enjoyment and efficiency in the learning process (Hsu, 2016; Sandberg et al., 2014). 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 effective game features were noted: Dourda, Bratitsis, Griva, and 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 and Devitt (2016) was the importance of goal orientation. Goal orientation, challenge, and collaboration all relate to a final feature: competition. Competition pushes players to set goals for winning, creates challenge, and also can build a sense of teamwork. However, similar to challenge, the effectiveness of competition is quite obscure. The mere mention of games evokes the idea of competition, yet it may not actually be beneficial when using games as learning tools. From the included studies, ten (38%) at least touch on competition, typically when sharing previous research or in describing the game used in their study. Five (19%) studies actually mention competition in their discussion or analysis, but what is interesting is they all make a similar point: while to some extent competition can act as a motivator, it also can distract the player from the learning, and reduce enjoyment or motivation (Butler et al., 2014; Cohen, 1960; Ebrahimzadeh & Alavi, 2016; Wong & Hsu, 2016; Yang et al., 2016). When designing *Chinese-PP* for their study, Hsu (2016) considered the issues surrounding competition and ensured a balance between collaboration and competition, where allies and enemies were never permanent.

These findings align with Boyle et al. (2016), who reported that rewards, cooperation, and goal-orientation result in positive affective and learning outcomes, while competition garnered mixed support. Thus, in designing a DLG, or using a DLG in the classroom, educators and designers should consider player control, challenge (in one's ZPD), instant feedback, mystery, collaboration, goal orientation, clear rules, and when appropriate, competition.

4.4. Outcomes from gameplay

Outcomes from gameplay were coded based on the research-supported categories and sub-categories used by Hung et al. (2018). They identified six potential outcome categories: *language acquisition, knowledge acquisition, contemporary competences, affective/psychological states, participatory behaviors,* and *correlational outcomes* (Hung et al., 2018). These categories are in line with the definition of outcomes presented in 2.1 in that they embody the various ways in which DGBLL can help a player grow and they are valued in many educational contexts. The coding of outcomes was based on the review by Hung et al. (2018), but there were slight alterations made depending on the content of the articles and focus of this paper. If it seemed there was an outcome that didn't fit into their coding, or one of their sub-categories wasn't covered in the articles, then a sub-category was added (through open-coding) or

excluded. For instance, L1 was excluded from language acquisition, while orthographic and phonological awareness were both added. Therefore, while their article can act as a reference to understand the research behind each category, the sub-categories are also described here.

In this review, there were ten sub-categories 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). Dourda et al. (2014) was the only article that reported on *knowledge acquisition* (subject-matter, cultural knowledge, or other types of knowledge gains), in this case, geography. This was a surprising finding given the importance of culture in the actual application of languages. Articles that reported on *contemporary competences* included to the sub-categories of 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* consider the relationship between individual characteristics and gameplay 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* or *negative* outcomes if the author reported significance in the results, with positive outcomes meaning DGBLL positively impacted the player in the given outcome category. 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 impact* means the authors reported no significant findings or differences between experimental 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 sub-category within a given category, while others discussed multiple sub-categories, meaning one category could have both *negative* and *positive* results. For example, under learning acquisition Sundqvist & Wikstrom (2015) reported outcomes related to vocabulary, listening, writing, and reading, some being positive and some being negative. Rather than reporting on each individual sub-category, all the outcomes 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 outcomes were 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.

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. This isn't surprising since the focus of this review is on language, but Hung et al. (2018) reported 37 language outcomes out of their 50 accepted articles, meaning 13 articles didn't discuss language outcomes at all. In both their review and the present study, vocabulary was the most commonly analyzed sub-category for language acquisition.

In comparing the results for each outcome, 70% of the 64 outcomes (excluding correlational) were found to be positive, similar to Hung et al. (2018) who reported 77% of the 105 outcomes from their included studies were positive. There was only one exclusively negative result (Hung et al. (2018) found none), reported by Cornillie et al. (2018) whose study enabled participants to practice as much as they wanted. By the end of the study, only a quarter of participants remained, but those who did continue to play had positive language acquisition outcomes. Other negative results were also found, as seen by the 10 (16%) mixed results, but those studies also reported positive results or no significant impact for at least one sub-category. For example, Salehi (2018) study had mixed language outcome results because he found no difference in initial vocabulary learning between groups after the experiment, but one month later found the experimental group (used DLGs) had higher vocabulary retention than the control group (used a booklet). 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 (drill practice) 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 game's impact (Hsu, 2016; Mazaji & Tabatabaei, 2016; Reitz et al., 2016).

Out of the correlational outcomes, four studies discussed gender differences, with two finding significant differences between genders. In the two studies with gender differences, it was clear that boys were more exposed to digital games than girls and thus

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

tended to do better in terms of related outcomes (Mazaji & Tabatabaei, 2016; Sundqvist & Wikstrom, 2015). Dalton and Devitt (2016) briefly discussed how gender might influence the type of games players enjoy, while Butler et al. (2014) found a gender effect in a couple of specific cases (game and level), but gender did not correlate with the frequency of gameplay. However, Butler et al. (2014) did find a correlation between age and frequency of gameplay: the frequency of plays decreased as age increased. Other correlations observed in studies include 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, 2016; Wong & Hsu, 2016). However, with only one study looking at each type of correlation, conclusions cannot be drawn.

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 competences, 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, but one can assume that DGBLL could be used to enhance cultural competences. 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, both concerning interaction and communication. Future research should explore the participant's choice to play, attrition rate, and how playing games impact the participant's activity in class as well. More detailed descriptions of the findings are provided in Appendix.

4.5. Study context

Another area of interest was the study context since the effectiveness of a DLG could differ depending on the environment in which it is played. There were three categories that articles could be coded into: *formal learning environment with educator facilitation* (FLF), *formal learning environment without educator facilitation* (FL), and *informal learning environment* (IL). In addition, 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 outlines the language acquisition result for each study in relation to the research context.

Twelve studies (46%) were conducted in a 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.

A FL is also an educational setting, but without any facilitation; the researcher or teacher may have given directions on how to play the game but the actual learning wasn't facilitated. 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 learning environment* (IL), which includes at-home use or a laboratory-type context. Out of the four studies, one yielded positive language acquisition results (25%), one had no significant impact (25%), and two had mixed results (50%). Given the low number of IL studies and mixed language outcomes, conclusions cannot be drawn.

Two experimental studies were found to take place in a combination of an FLF and an IL (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 et al., 2018). In other words, users could choose whether and when they wanted to play, and so over time many stopped playing altogether. Since students did not explain why they stopped playing, this finding lacks weight.

5. Discussion

The aim of this review was to explore the available empirical evidence for DGBLL, and to better understand how DLGs can be used to improve outcomes by considering game features and context. Previous researchers have emphasized the importance of the teacher's facilitation of the game in order for it to be most effective (Beavis 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). In total, the 18 studies within a FL and FLF were 78% successful in improving language skills, and thus it can be claimed that DLGs are

Table 2
Study context and the results.

	Formal learning environment with educator facilitation	Formal learning environment without educator facilitation	Informal learning 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

effective tools for teachers, with or without facilitation. However, none of the studies directly looked at the context of the study and effective teacher facilitation. Through an experiment, future researchers could examine effective game facilitation and explore specific teaching strategies that impact outcomes. This would be an answer to an issue raised by Wastiau et al. (2009): 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 classroom time. In addition, 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 (Wastiau et al., 2009). Therefore, more research should reveal tangible ways teachers can successfully and easily use DLGs.

Context matters in education since learners need a safe environment where they are willing to interact, which DLGs provide (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, 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 players' intrinsic motivation and feelings of autonomy (Hamari et al., 2016; Peterson, 2010; Squire, 2008; Wilkinson, 2016), both of which were supported in this review. Surprisingly, none of the included articles had games that enhanced cultural competence or allowed for intercultural communication. The internet is an ideal environment to connect users internationally, yet these studies failed to do so. Despite the strong evidence in support of interactivity, more research should look at how DLGs can be used to improve sociocultural outcomes.

In addition to context, the features of the DLGs also influence outcomes. Based on the articles used in this review and findings reported by Boyle et al. (2016), 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 player interaction (Butler et al., 2014; Dalton and Devitt, 2016; Dourda et al., 2014; Ebrahimzadeh & Alavi, 2016; Hwang et al., 2016; Wu et al., 2014; Yang et al., 2016). Sandberg et al. (2014) compiled a list of effective game features that they implemented in their game design, but they weren't directly tested: fantasy, rules and goals, sensory stimuli, challenge, mystery, control, rewards, and competition or cooperation. With the exception of "fantasy", they are nearly identical to the ones found in this review. Of course, individual 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 and so with a game structured around these features, players should be more successful.

With that said, features may impact one outcome category and not another, as revealed by the mixed results in some studies. For instance, Butler et al. (2014) found certain games attracted players but didn't necessarily enhance language skills. Furthermore, the level of challenge and ability for players to communicate in the game *Haunted* increased motivation and communication skills, but players were typically unaware of language mistakes and the game completion didn't depend on language quality (Reitz et al., 2016). Competition was the most controversial feature discussed as it could at times enhance motivation, but it also was found to distract the player from the learning or even demotivate players (Butler et al., 2014; Cornillie et al., 2018; Ebrahimzadeh & Alavi, 2016; Wong & Hsu, 2016; Yang et al., 2016). It can thus be said that researchers should further investigate specific features that cause users to become distracted and detract from the learning. This is a clear limitation to the current review, as it is not possible to explicitly define which features will enhance outcomes in all categories, and all of the authors were focused on the positive aspects of DGBLL.

Considering the lack of research on negative outcomes, publication bias must be taken into consideration. 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, which could inform the DLG world about features that cause negative outcomes.

A final limitation of the studies included in this review is the lack of longitudinal research. While there were three longitudinal studies that looked at retention, results were mixed and the maximum length of time was two months (Salehi, 2018; 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, with differing game genres, and for a range of participants. Further research into what makes certain games more effective than others should be explored.

6. Conclusion

This review analyzed 26 studies looking into the impact of digital games on language learning and related outcomes from 2014 to 2018. The majority of research was mixed methods, targeting English language learning using computers, and was conducted in East Asia or the Middle East. It was found that DLGs can be used as effective L2 learning tools that motivate players to learn and interact. DGBLL can be a fun, engaging, and challenging way to learn, and provides differentiation and learner autonomy. From the included studies, 70% of the reported outcomes were entirely positive. This is evidence of the positive outcomes of DGBLL on primary through high school-age children. More specifically, researchers reported outcomes from DLGs were positive 62% of the time for language acquisition, 81% for affective/psychological states, 88% for contemporary competences and 62% for participatory behaviors. While the articles did not mention outcomes related to cultural competences or building global networks, DLGs provide the possibility of expanding the classroom outside of four walls by bridging schools and enabling cross-cultural communication.

In order for DLGs to be implemented successfully, it is essential to know how certain factors can influence the outcomes. This review found DLGs produced positive outcomes with and without teacher facilitation, but more insight into how teachers can implement DGBLL is needed since none of the studies directly analyzed implementation. While there were no clear associations

between gaming context and impact, specific game features were noted across articles that are thought to improve outcomes: ease-of-use, meeting one's ZPD, rewards and feedback, control or autonomy, goal-orientation, and interactivity. Competition as a feature was found to both enhance and reduce outcomes, and thus further research should more closely examine which features negatively impact players.

Digital devices are part of the NetGen'ers everyday life and can be used either as tools or toys. DLGs act as both, and it would benefit educators to know how to successfully use them to improve the learning experience. By considering the aforementioned features to design games specifically for use in classrooms, and by creating practices for teachers to successfully implement them, DLGs may revolutionize the learning environment. Future research should specifically explore DGBLL for different age groups and how teachers at the primary, secondary and even university level may incorporate DLGs into their teaching.

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 behavior-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.

(continued on next page)

Table A1 (*continued*)

Article	Aims	Methods* Game details**	Outcomes
(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 = 40 (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, 2018)	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.
(Shokri & Abdolmanafi-Rokni, 2014)	To investigate the effectiveness of using computer games on recall and retention of word spelling in Iranian EFL male and female students.	N = 40 (14–15); FLF; English; Quan; spelling pre- and post-tests GL, Comp, MG: Hangman, Spelling bee, Fast hand, and Concentration Games	There was a significant difference between groups in recall of spelling, with the 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, Comp, 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.

FL = formal learning environment without game facilitation; FLF = formal learning environment with game facilitation; IL = informal learning environment; Quan = quantitative; Qual = qualitative.
GL = game for learning; EG = entertainment game; Comp = Computer; Ed = educational; MG = educational mini-games; RPG = role-play; Ad = adventure; St = strategy; Sim = simulation.
*Methods: number of participants (age); context; target language; data type; any additional information about the research design and measures.
**Game details: game purpose; platform used; game genre: name of game.

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 gameplay 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.

Table A2 (*continued*)

Article	Aims	Methods* / Game details**	Outcomes
(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 website with 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 (designed specifically for the purpose of the research)	The digital game enhanced the vocabulary of the participants with the experimental group (gameplay) 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, Ed: N/A (researchers developed a problem-based English listening game using RPG maker)	The experimental group (gameplay) performed significantly better than the control group (traditional teaching) in English listening (F = 17.53, p < 0.001, n² = 0.19). The same can be said for learning motivation (F = 26.24, p < 0.001, n² = 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. GL, Comp, MG: (researchers developed a digital game with the badge mechanisms)	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) 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

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

GL = game for learning; EG = entertainment game; Comp = Computer; Ed = educational; MG = educational mini-games; RPG = role-play; Ad = adventure; St = strategy; Sim = simulation.

*Methods: number of participants (age); context; target language; data type; any additional information about the research design and measures.
**Game details: game purpose; platform used; game genre: name of game.

Table A3

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

Article	Aims	Methods* Game details**	Outcomes
(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 = 3945 (4–12); IL; English; Quan, Correlational; descriptives of game use, and mock test and Jido-Eiken test scores to measure English performance. GL, Comp, MG: Jido-Eiken learning game	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.
(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. EG, Comp, Sim: Open Sim (researchers developed a private, secure virtual world)	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.
(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. EG, Comp, Ad: Whodunit	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 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 behavior 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 & Wikström, 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

Table A3 (*continued*)

Article	Aims	Methods*	Outcomes
		Game details**	
			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, Ad: 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 (researcher developed)	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).

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

GL = game for learning; EG = entertainment game; Comp = Computer; Ed = educational; MG = educational mini-games; RPG = role-play; Ad = adventure; St = strategy; Sim = simulation.

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

**Game details: game purpose; platform used; game genre: name of game.

References

Coded Papers

deHaan, J. (2011). Teaching and learning English through digital game projects. *Digital Culture & Education, 3*(1), 46–55. http://www.digitalcultureandeducation.com/cms/wp-content/uploads/2011/04/dce1046_dehaan_2011.pdf.

Abdullah, S., & Akhter, J. (2015). Ego is a hurdle in second language learning: A contrastive study between adults and children. *Advances in Language and Literary Studies, 6*(6), 170–173. https://doi.org/10.7575/aiac.alls.v.6n.6p.170.

Aghlara, L., & Tamjid, N. H. (2011). The effect of digital games on Iranian children's vocabulary retention in foreign language acquisition. *Procedia - Social and Behavioral Sciences, 29* http://doi.org/10.1016/j.sbspro.2011.11.275.

All, A. (2014). Measuring effectiveness in digital game-based learning: A methodological review. *International Journal of Serious Games, 1*(2) https://doi.org/10.17083/ijsg.v2i4.98.

Ashraf, H., Motlagh, F. G., & Salami, M. (2014). The impact of online games on learning English vocabulary by Iranian (low-intermediate) EFL learners. *Procedia - Social and Behavioral Sciences, 98*, 286–291. http://doi.org/10.1016/j.sbspro.2014.03.418.

Beavis, C., Rowan, L., Dezuanni, M., McGillivray, C., O'Mara, J., Prestridge, S., et al. (2014). Teachers' beliefs about the possibilities and limitations of digital games in classrooms. *E-Learning and Digital Media, 11*(6), 569–581. http://10.2304/elea.2014.11.6. 569.

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open, 2*, 8–14. https://doi.org/10.1016/j.npls.2016.01.001.

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., 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. 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 a foreign language setting: A meta-analysis. *British Journal of Educational Technology, 43*(4), E104–E107. https://doi.org/10.1111/j.1467-8535. 2012.0129 5.x.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.

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. https://doi.org/10.1016/j.compedu.2012.03.004.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience* (1st ed.). New York: Harper & Row.

Gee, J. (2004). *Situated language and learning: A critique of traditional school.* New York: Routledge.

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. 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. https://doi.org/10.1016/j.chb.2015.07.045.

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition, 177*, 263–277. https://doi.org/10.1016/j.cognition.2018.04.007.

Herz, J. C. (1997). *Joystick nation: How videogames gobbled our money, won our hearts, and rewired our minds* (1st ed.). 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. 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), E6–E10. https://doi.org/10.1111/j.1467-8535.2011.01242.x.

Kim, Y. (2018). Analysis of research trends focusing on outcomes in language learning using MMORPGs. *Multimedia-Assisted Language Learning, 21*(4), 111–142. Retrieved from http://journal.kamall.or.kr/wp-content/uploads/2019/01/5Yunjin-Kim.pdf.

Klimova, B., & Kacet, J. (2017). Efficacy of computer games on language learning. *Turkish Online Journal of Educational Technology, 16*(4) https://eric.ed.gov/?id=

EJ1160637.

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

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. 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. https://doi.org/10.1371/journal.pmed.1000097.

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

Newhouse, C. P. (2014). Learning with portable digital devices in Australian schools: 20 years on!. *Australian Educational Researcher, 41*(4), 471–483. https://dx.doi.org/10.1007/s13384 -013-0139-3.s.

Noddings, N. (2015). *Philosophy of education.* Boulder: Routledge.

Oblinger, D. (2004). The next generation of educational engagement. *Journal of Interactive Media in Education, 1*, 1–18. 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. 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)https://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. 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. https://doi.org/10.1080/09588221. 2016.1197949.

Potter, W. J., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in content analysis. *Journal of Applied Communication Research, 27*(3), 258–284. https://doi.org/10.1080/00909889909365539.

Prensky, M. (2001). The digital game-based learning revolution. *Anonymous digital game-based learning.* United States: McGraw-Hill Publishing.

Roschelle, J. M., Pea, R. D., Hoadley, C. M., Gordin, D. N., & Means, B. M. (2000). Changing how and what children learn in school with computer-based technologies. *The Future of Children, 10*(2)https://doi.org/10.2307/1602690.

Sang, Y. (2017). A conceptual review of age effect on L2 acquisition. *Journal of Education and Practice, 8*(9), 1–4. Retrieved from https://files.eric.ed.gov/fulltext/EJ1139050.pdf.

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. 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), Retrieved from http://www.gamestudies.org/0102/squire/.

Squire, K. D. (2008). Video games and education: Designing learning systems for an interactive age. *Educational Technology, 48*(2), 17–26. Retrieved from https://website.education.wisc.edu/kdsquire/tenure-files/02-squire-ed-tech-refchecV3.pdf.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Wastiau, P., Kearney, C., & Van den Berghe, W. (2009). *How are digital games used in school?* European Schoolnet. Retrieved from http://games.eun.org/upload/gis-full_report_en.pdf.

Wilkinson, P. (2016). A brief history of serious games. In R. Dörner,, S. Göbel,, M. Kickmeier-Rust,, M. Masuch,, K. A. Zweig,, & R. Dörner, (Eds.). *Entertainment computing and serious games* (pp. 17–41). Cham: Springer.

Young, M.,F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., 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. https://doi.org/10.3102/003465431243 6980.

Yudintseva, A. (2015). Synthesis of research on video games for the four second language skills and vocabulary practice. *Open Journal of Social Sciences, 3*, 81–98. http://doi.org/10.4236/jss. 2015.311011.

Zuengler, J., & Miller, E. R. (2006). Cognitive and sociocultural perspectives: Two parallel SLA worlds? *Tesol Quarterly, 40*(1), 35–58. http://10.2307/40264510.

Butler, Y. G., Someya, Y., & Fukuhara, E. (2014). Online games for young learners' foreign language learning. *English Language Teaching Journal, 68*(3), 265–275. 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 and Technology, 23*(3), 10–26.

Dalton, G., & Devitt, A. (2016). Action research Irish in a 3D world: Engaging primary school children. *Language, Learning and Technology, 20*(1), 21–33. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44440/1/20_01_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. Retrieved from www.ejel.org/issue/download.html?idArticle=285.

Ebrahimzadeh, M. (2017). Readers, players, and watchers: EFL students' vocabulary acquisition through digital video games. *English Language Teaching, 10*(2), 1–18. https://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. https://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. 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. http://doi.org/10.1080/09588221.2017.1329215.

Hsu, T. (2016). Learning English with augmented reality: Do learning styles matter? *Computers & Education, 106*, 137–149. http://doi.org/10.1016/j.compedu.2016.12.007.

Hwang, J., 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. https://doi.org/10.1016/j.compedu. 2016.11.010.

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. https://doi.org/10.1080/09588221.2015.1016438.

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. *Computers & Education, 102*, 188–201. http://doi.org/10.1016/j.compedu.2016. 07.005.

Lan, Y. J. (2015). Contextual EFL learning in a 3D virtual environment. *Language, Learning and Technology, 19*(2), 16–31. Retrieved from https://www.researchgate.net/publication/2810269 29_Contextual_EFL_learning_in_a_3D_virtual_environment.

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. Retrieved from http://mjltm.org/browse.php?mag_id=24&slc_%20 lang=en&sid=1.

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. 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. https://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. https://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. 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. 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. 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. 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. Retrieved from 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. 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 & Development, 62*, 209–226. 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. 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. Retrieved from https://eric.ed.gov/?id=EJ1039075.