

The effectiveness of Game-Based Science Learning (GBSL) to improve students' academic achievement: A meta-analysis of current research from 2010 to 2017

*¹Heru Setiawan; ²Shane Phillipson

¹Sekolah Menengah Atas Global Mandiri Jakarta

Jl. Raya Cakung Cilincing KM. 5, Cakung Timur, Jakarta Timur, DKI Jakarta 13910, Indonesia

²Faculty of Education, Monash University, Australia

29 Ancora Imparo Way, Clayton VIC 3800, Australia

*Corresponding Author. E-mail: heru.setiawan@teacher.globalmandiri.sch.id

Submitted: 10 November 2019 | Revised: 6 December 2019 | Accepted: 12 December 2019

Abstract

This study identifies the effectiveness of game-based science learning (GBSL) for improving students' learning outcomes by conducting a literature review of the current research from 2010 to 2017. This study also explores the correlation between variation in school level and year of publication on GBSL effect size. Data were collected from peer-reviewed journal articles published in educational databases including ERIC (Educational Research Information Centre), Springer Link, ProQuest education journal, and A+ education. Seven inclusion criteria were used to select relevant studies. Comprehensive Meta-Analysis (CMA 2.0) was used to analyze the data. This study finds that (1) GBSL intervention has a statistically significant effect on students' learning outcomes with a higher average on the effect size of the experimental group (41.12) than the control group (37.07). The mean of the reviewed studies' effect size is 0.667 in the medium category. (2) The implementation of GBSL in secondary school has a bigger average effect size than in elementary school. Year of publication and effect size has a low positive correlation with a coefficient of correlation 0.40.

Keywords: *game-based science learning, learning outcomes, meta-analysis*

Permalink/DOI: <https://doi.org/10.21831/reid.v5i2.28073>

Introduction

The young generation who was born in the 21st century is a digital native or the Net generation (Bennett, Maton, & Kervin, 2008). The millennial in this era also can be called a game generation (Prensky, 2001). The trend of digital games' use has been increasing in this era (Corbett, 2010; McGonigal, 2011). Millions of people have been immersed in playing digital games either for entertainment or education (Huang, Hew, & Lo, 2019). Gee (2007) reported in his study that approximately 90% of students' mobile phones connect to digital games. Besides, many teachers use digital games as a medium of instruction in their classroom for engaging students dur-

ing teaching and learning processes, or it is commonly called digital game-based learning (DGBL) (Papastergiou, 2009; van Eck, 2006). Students also obtain feedbacks such as improvement, and win conditions after completing the goals (Okeke, 2016, p. 1). The DGBL that specifically focus on Science is called Game-Based Science Learning (GBSL).

Since 2006, the number of research investigating the effect of Digital games in education has been increasing (Chorney, 2012). Some literature has been debating the effectiveness of GBSL in the last decade (Hamari & Keronen, 2017; Quandt et al., 2015). The community of science education (physics, biology, chemistry, and general sciences) also

concern with the potential of game-based learning. Some researchers investigate the effectiveness of GBSL in some science subject matter such as Newtonian mechanics (Clark et al., 2011), human immunology (Cheng, Su, Huang, & Chen, 2014), and photosynthesis (Culp, Martin, Clements, & Lewis Presser, 2015). They argue that science is challenging for some students because of abstract concepts and invisible objects. In addition, some research illustrated that rote memorization and decontextualized learning have potential drawbacks in the Science context (Honey & Hilton, 2011; Mayo, 2007). This issue has an impact on their learning outcomes which can be defined as skills, knowledge, and values as an outcome of students' experiences (The US Council for Higher Education Accreditation (CHEA) cited in Adam (2004, p. 4). Learning outcomes can be knowledge, skills, or attitude. However, in this context, the learning outcome only refers to students' learning outcomes in academic settings. Thus, GBSL is the proper solution to this issue because digital games are highly engaging and motivating (Huang et al., 2019; Tsay, Kofinas, & Luo, 2018). Several researchers demonstrated empirical evidence of the potential of this educational tool to enhance students' learning outcomes in the various context of science subjects through comparing control and experiment group (such as Bello, Ibi, & Bukar, 2016; Fan, Xiao, & Su, 2015).

However, a small number of sample of studies investigating the effect of GBSL on students' learning outcomes tended to have a more significant mean of effect sizes than studies with larger sample sizes (Cheung & Slavin, 2013). Effect size refers to a quantitative measurement of the difference between the mean score of the control group and the treatment group (Nakagawa & Cuthill, 2007). Meanwhile, the small sample size of the research cannot be used to generalize the effect of GBSL. In order to solve this issue, it needs further investigation of the effectiveness of GBSL in students' achievement in sciences with a meta-analysis study to develop a better estimate of effect magnitude (King & He, 2005). Meta-analysis is the process of converting the effects of several similar research

into quantitative data so that these averages of the effect size and an overall determination can be made concerning the cumulative findings of several studies (Glass, McGaw, & Smith, 1981). Meta-analysis is a kind of retrospective observational study in which researchers make data recapitulation without any experimental manipulation (Brockwell & Gordon, 2001).

Several literature reviews of Game-Based Learning have been conducted both in the context of sciences and other subjects such as mathematics, language, history, and physical education. In 2006, Vogel et al. (2006) used meta-analysis of digital games versus traditional teaching methods. The overall result of the meta-analysis was that treatment groups were reported higher learning outcomes and better attitudes toward learning than control groups. The report also analyses some moderator categories. He reported that gender, school level, and user type showed significant statistical results. Meanwhile, learner control, type of activity, and realism do not appear to be influential. In the science context, Li and Tsai (2013), reviewed research articles regarding game-based science learning (GBSL) published from 2000 to 2011. The focus of the review is qualitative outcomes including research purposes and designs, the theoretical foundations, game design, and learning focus. Based on the review, GBSL can provide effective learning in a collaborative problem-solving environment. However, the research only focused on qualitative data without discussing and analyzing the quantitative analysis of GBSL intervention and the effect size.

According to the previous research, gaps in the literature have been identified. Although several studies have explored a review of literature of GBSL, few have tested their relative influence on learning the outcome. There was also a lack of research meta-analysis of GBSL with a quantitative approach. Li and Tsai (2013) who focused their research on the qualitative method suggested that quantitative content analysis of GBSL effectiveness such as students' learning outcomes in Science education should be conducted in future investigations. It is because

digital games that can promote students' engagement (Annetta, Minogue, Holmes, & Cheng, 2009; Tsay et al., 2018) might also enhance students' learning outcomes. Other similar studies such as Vogel et al. (2006) also have a limitation. Although he specifically focusses on cognitive aspects in the analysis, the context of the study is in a broad context and did not specifically focus on Science education. Based on this gap, a newly proposed work focusing on a meta-analysis of the effect of the digital game on students' learning outcome in Science education or GBSL need to be conducted. Thus, two central research questions (RQs) were addressed in this study: (1) RQ1: Do Game-Based Science Learning (GBSL) effective to enhance students' learning outcomes compared to traditional method as reported by the current studies from 2010 to 2017? (2) RQ2: Do moderator categories including school level of participants (elementary and secondary school context) and year of publication has any correlation with GBSL effect size?

This research contributes to the literature in this field. First, this study reviewed recent trends in GBSL research, especially for those in the field of science education who are interested in quantitative studies of GBSL for students' learning outcomes. A meta-analysis of GBSL has been conducted by several researchers within a broader context such as mathematics, language, and other subjects (Divjak & Tomić, 2011; Young et al., 2012), but lack of research conducted in science education. Second, the consistency of the result of similar studies for several years will be investigated. Therefore, consistency and inconsistency of findings of similar research will be found, and bias of one or more studies in this field could be detected (Borg & Gall, 1983). Third, a meta-analysis uses a significant amount of data, and applying statistical methods by organizing some information comes from a broad cross-section whose function is to complement other purposes (Glass et al., 1981). By the significant number of participants, the study develops a better estimate of effect magnitude (King & He, 2005). The larger sample size in conducting a meta-analysis could be found in one study that will

create greater statistical power and more precise confidence intervals. This is because the study collects several similar studies to be analyzed quantitatively. It concentrates on the effect size of this empirical discovery which is relatively better than the other methods of quantitative approaches including narrative review, descriptive review, and vote counting (Lipsey & Wilson, 2001). Moreover, through the substantial number of participants with different variables, the differences may exist because of differences that exist among the articles such as different subject populations, education level, gender, game type, etc. By using meta-analysis, different moderator variables can be investigated. Vogel et al. (2006) state that analyzing moderator variables would give a clearer overview or more complex picture of reviewed studies.

Method

Research Strategies and Data Collection

The search of the literature was conducted from June to July 2017. Data were collected from journal articles published from educational sources including ProQuest education journal, Springer Link, A+ education, and ERIC (Educational Research Information Centre). The databases provide a high impact and a high-quality journal article. The keywords are 'digital game, sciences, physics, biology, chemistry, secondary, high school, elementary.' The Boolean operator, 'AND' or 'OR', was used to combine all key terms. Following the keywords, the researchers read the abstract and full-text. We use some inclusion and exclusion criteria as the evaluation to choose appropriate journal articles. Seven inclusion and exclusion criteria were applied in screening the eligible article included in this study including publication year, unit, participant, game intervention, research design, participant, outcome type, and language. These details of inclusions and exclusions are explained as follows.

(1) Publication year: All of the articles are peer-reviewed journal articles published in the last seven years from January 2010 to June 2017. (2) Unit: The unit in elementary and secondary education in this study is science

subjects including biology, physics, chemistry, and general sciences. Other units such as technical subjects in vocational high school are excluded. Also, unrelated subject matters that have similar keywords, but they are not related to science subjects such as physical education are excluded. (3) Game/ intervention: Digital games in this study is defined as a digital experience where participant use game of computer software and they receive feedback to achieve the goals in the form of a score, progress and win condition. However, learning intervention that focused on creating a digital game for students is not included. The studies compared digital games in science instruction and traditional methods. (4) Research design: All of the journal articles included in this meta-analysis must use experimental and control groups or game versus non-game conditions. The studies must have a sample size, standard deviation, and mean. However, studies that do not have the data were excluded. The studies included used an experimental method to make sure that the included studies have data compared in the statistical analysis. Studies are considered experimental if individual students are randomly assigned to an instructional condition. (5) Participant: The participants of the research in the included studies are elementary and secondary school students. Students with specific clinical criteria such as disabilities are excluded from this study. (6) Outcome type: The data that will be extracted in this study is only quantitative data (numerical data) specifically students' learning outcome or cognitive aspect. Other research outcomes or qualitative data such as behavior, activity, participation, collaboration, engagement, and motivation are not extracted. (7) Language: The study included is an only article published in English without considering the country in which the studies are conducted.

The full text that is related to the inclusion criteria of the topic was evaluated by annotating each article to extract some necessary information. This step was conducted using note-card contained eligibility criteria evaluation rubric recommended by Mertens (2015) including research question, the design of research, data analysis, results, conclusion,

and research evaluation. During the preliminary selection of eligibility occurred in 137 articles were identified. Then, after the articles were screened for eligibility to exclude some non-eligible full text by applying inclusion criteria, 12 journal articles are carefully selected although this amount is a small number relative to some meta-analyses in this field.

The data from the selected studies is then extracted for further analysis. First, the data of the characteristics of the reviewed studies that include the year of publication, country of origin, school level of participants, science domain, game name, and the purpose of the study were noted in Microsoft Excel. The data were extracted through manual searches in each article. The data is important to provide an overview of the characteristics of the reviewed studies. Second, the key information which corresponds to the research questions were also extracted for each study. The information which is needed to answer the research questions is only quantitative data (numerical data) that was used in the statistical analysis. The quantitative data extracted are student's achievement means, standard deviation, the number of participants of the control and treatment group.

Data Analysis Method

Microsoft Excel and Comprehensive Meta-Analysis (CMA 2.0) were used for statistical analysis after the quantitative data were extracted. Formerly, the demographic characteristics of the reviewed studies were analyzed with descriptive statistics using Microsoft Excel which present data such as mean, percentage, and also frequencies. The data would also be presented with visual techniques such as a column, bar chart, and histogram. Lately, CMA 2.0 was also used. Several researchers verified the accuracy of the analysis method (Ones, Viswesvaran, & Schmidt, 1993). CMA 2.0 is used to analyze Hedges' *g* effect size, the lower limit (LL), the upper limit (UL), *p*-Value, and the Relative weight of all studies (Borenstein, Hedges, Higgins, & Rothstein, 2005). In order to give a clearer overview of the overall effect size, the forest plot to compare the effect of digital games over traditional methods was used (Sutton, Abrams,

Jones, Sheldon, & Song, 2000). Two kinds of effect models in a meta-analysis are fixed effect model and random effect model (Michael Borenstein, Hedges, Higgins, & Rothstein, 2010). The decision to select the effect model to analyze data is an essential factor in the meta-analysis (Hedges & Vevea, 1998). Improper determination of the model will cause inefficient estimation and incorrect conclusions (Nickell, 1981). However, in this study, we use the random effect model because all twelve studies which are used in this research were drawn from different populations, such as different populations in different countries. A similar condition of research is conducted by Sacks, Berrier, Reitman, Ancona-Berk, and Chalmers (1998). Moreover, the studies report varies the effect size (ES). In the random-effects model, the true effect size might differ from one study to another study (Olejnik & Algina, 2000). In addition to the estimation of the primary effect, secondary analyses were conducted to take advantage of the coded study characteristics and test the moderating effects. Specifically, secondary analysis tested the influence of grade level (elementary and secondary school) and year of publication. The data from statistical analysis from CMA 2.0 were used in order to address the research questions with the following method of interpretation.

We address the first research question by comparing the experimental group and the control group. There would be no difference between the control and experimental group when the mean of the sample is equal. However, when the experimental group's means score is higher than the control group, it means that GBSL intervention is more effective through looking at the mean difference between the experiment and the control group. The second research question is answered by investigating the effect of moderator categories including the year and school level, to the GBSL effectiveness, we use descriptive analysis by comparing the mean of effect size in each category. We compare the average effect size at each school level (elementary and secondary school) to determine which school level more effective in the game intervention. Then, to analyze whether or not

publication year has any correlation with game effectiveness, we use inferential statistics because it strives to make inferences and predictions (Bryman, 2016). The statistical method would improve the previous research that only looks at the pattern of effect size across the years. The data would be presented as scatterplot to illustrate the relationship between two variables (Cohen, Manion, & Morrison, 2007, p. 507). It would also count the Spearman's rank correlation coefficient (r) because both variables are ordinal to see the linear trend using Microsoft Excel. The interpretation to assess the degree of the correlation coefficient were categorized into very high (0.9 to 1.0), high (0.7 to 0.9), moderate (0.5 to 0.3), low (0.3 to 0.5), and negligible correlation (0 to 0.3).

Detection of Publication Bias

Detection of publication bias of reviewed studies is crucial in meta-analysis study (Rothstein, Sutton, & Borenstein, 2006). Publication bias is the tendency of researchers to screen articles for publication based on the statistical significance of effects than the quality of the study (Rothstein et al., 2006, p. 296). Several pieces of evidence show that some research that has a higher effect size is more likely to be published (Peters, Sutton, Jones, Abrams, & Rushton, 2006). Consequently, it will affect the review process. Therefore, the meta-analysis may be overestimated effect size because it uses a biased sample or target of the population. Hence, to avoid this concern or minimizing this bias in this study, it needed a model to know which study is missing. One of the proper models is the funnel plot (J. A. C. Sterne et al., 2011). In the funnel plot, the effect size is plotted in X-axis, and the number of participants is plotted in Y-axis (Sterne & Egger, 2001). Also, asymmetry easily detected in the funnel plot. The studies will be distributed symmetrically when the publication bias is absent (Schmidt & Hunter, 2014). The next problem is whether the observed overall effect is robust. To solve this issue, some researchers use Rosenthal's Fail-safe N. Orwin (1983) suggested that Rosenthal's Fail-safe N compute the number of studies that should be incorporated in the analysis.

Findings and Discussion

Overview of the Reviewed Studies

The publication years range from 2010 to 2017. The purpose is to know the development of research in this area in the last eight years. The highest number of publications is in 2015 with three publications (Figure 1). Then, the presence of international studies is reflected in the sample. However, 50% of the studies included were conducted within the Asia continent especially in Taiwan, while the others were conducted internationally. There are two countries including Taiwan and Singapore from Asia. Within this interna-

tional group, Spain is well represented by two studies, while the other research is from the U.S and Nigeria, Africa (Figure 2). Based on the school level, elementary and secondary education has an almost equal number. Eight studies are from elementary school and four studies from high school (Figure 3). Subject areas are also well represented with three in the context of biology, seven general sciences, while each physics and chemistry are only one study (Figure 4).

The studies included are presented in Table 1. Table 1 outlines the characteristics of the included studies meeting all the eligibility criteria.

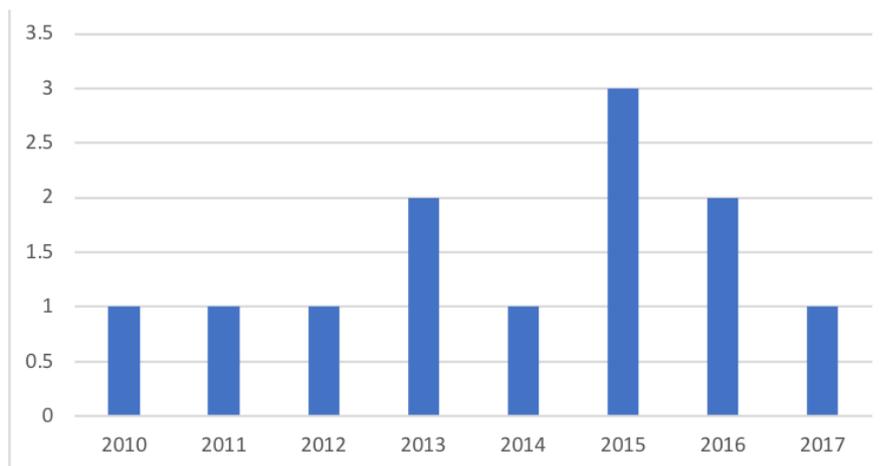


Figure 1. The number of reviewed studies by year of publication

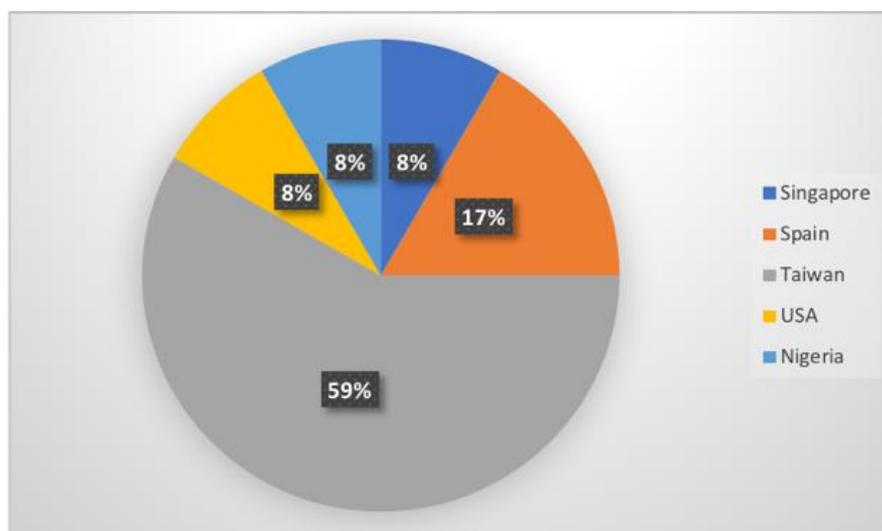


Figure 2. The reviewed studies by country

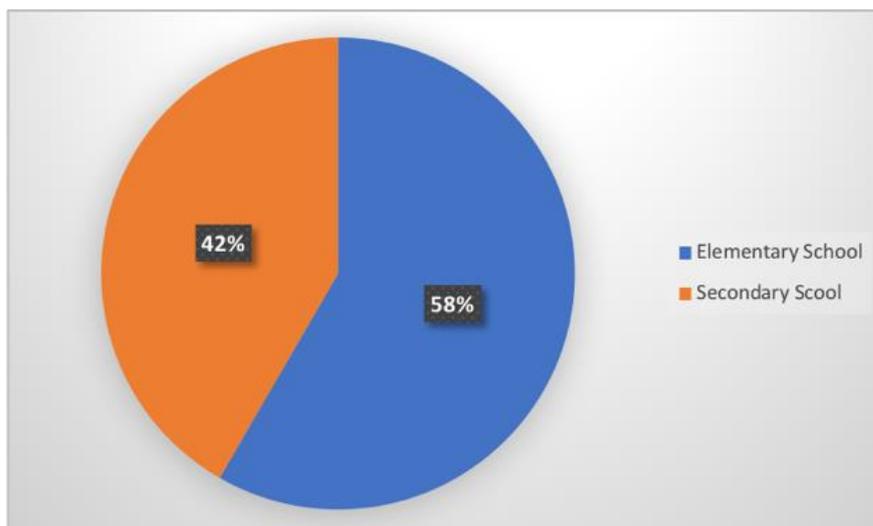


Figure 3. The reviewed studies based on the education level of participants

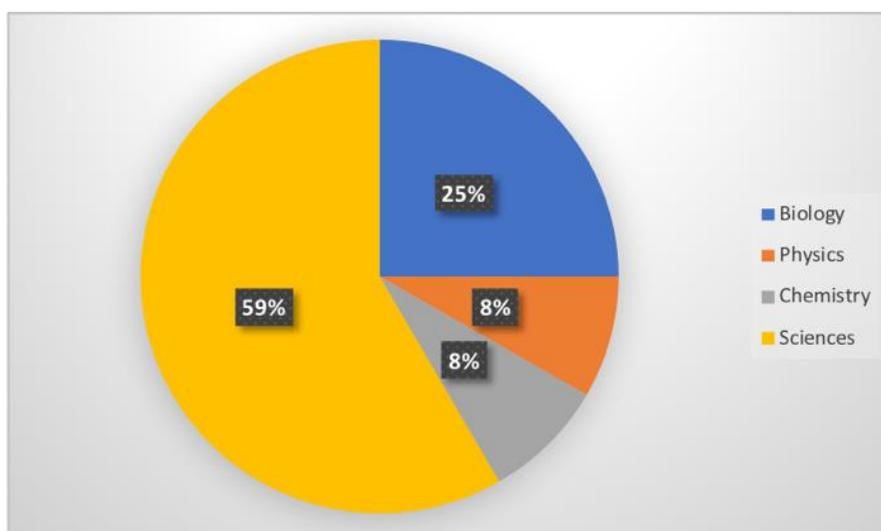


Figure 4. The reviewed studies according to science domain

Table 1. The background information of reviewed articles

Authors	Country	Game Name	School Level	Science Domain
Bello et al. (2016)	Nigeria	n/a	Secondary	Sciences
Chee & Tan (2012)	Singapore	Alkhimia	Secondary	Chemistry
Wrzesien & Raya (2010)	Spain	Supercharged	Elementary	Sciences
Anderson & Barnett (2013)	USA	Supercharged	Secondary	Physics
Sung & Hwang (2013)	Taiwan	Alien Invasion	Elementary	Sciences
Yien, Hung, Hwang, & Lin (2011)	Taiwan	Nutrition Supplement Battle	Elementary	Biology
Chu & Hung (2015)	Taiwan	Kodu	Elementary	Sciences
Su & Cheng (2015)	Taiwan	Find Insect	Elementary	Sciences
Chen & Hwang (2017)	Taiwan	Alien Invasion	Elementary	Sciences
Fan et al. (2015)	Taiwan	The MMBCLS	Secondary	Biology
Furió, Juan, Seguí, & Vivó (2015)	Spain	iPhone game	Elementary	Sciences
Chen, Yeh, & Chang (2016)	Taiwan	Role Play Game (RPG)	Secondary	Biology

How Effective GBSL Does to Enhance Students' Learning Outcomes in Sciences Compared to the Traditional Method as Reported by the Current Studies from 2010 to 2017?

The first research question is answered by comparing the average Mean of the reviewed studies. The result of data extraction is presented in Table 1 which compares the twelve studies with the treatment group and control group. The number of participants in the twelve studies is 954 students. Most of the studies have an equal number of participants in the treatment and control group, although some of them have a slightly higher participant in one group than the other group. There are 489 students in a total of the control group and 465 students from the experimental group. The number of participants in the studies is varied from 38 to 180 students. The standard deviation of all of the studies is also varied from the lowest 0.93 to the highest 23.54. The detail of the data for each study is shown in Table 2.

Based on Table 2, the average learning outcome mean from the overall studies of the experimental group (40.82) is higher than the control group (36.82). The mean difference analysis shows that one study, Chu and Hung (2015), has a negative mean difference between experimental and control group compared to the other ten studies that have a positive

mean difference. The highest mean difference between the studies is 19.63, while the lowest mean difference is -15.03. The experimental and control group's standard deviation shows a variation.

The Analysis Result of Standardized Mean Difference Effect Size, Variance, Weight, and Confidence Interval (CI)

The random-effects model was used to know the composite effect size with Comprehensive Meta-Analysis (CMA). The summary of the final analysis for all studies is presented in Table 3. We calculate Hedges's *g* for each study separately to maintain consistency of measurement. In addition to the individual effects, we also present a 95% confidence interval (lower limit and upper limit) around each study and the relative weight (*W*). The overall effect size of the twenty studies is $g = 0.661$, $p < .001$; with a 95% confidence interval between 0.223 and 1.090. It indicates a moderate overall effect for the synthesized GBSL interventions that is statistically different from a null effect. The largest effect size influencing this study is Bello et al. (2016) of 2.338. In contrast, the study contributing the smallest overall influence is Chu and Hung (2015) with an effect size of -0.637. The comparison of the SMD effect size of all studies is presented in a forest plot in Figure 5.

Table 2. Mean, standard deviation, and sample size of the studies on digital games versus control method

Authors (year)	Experiment Class			Control Class			N Total	Mean Difference
	Mean	SD	N	Mean	SD	N		
Bello et al. (2016)	66.23	7.07	90	46.6	9.48	90	180	19.63
Chee & Tan (2012)	3.28	2.61	40	2	1.71	38	78	1.28
Wrzesien & Raya (2010)	6.33	2.2	24	5.88	1.54	24	48	0.45
Anderson & Barnett (2013)	6.3	1.2	32	5.9	1.27	32	136	0.4
Sung & Hwang (2013)	57.26	16.87	31	43.07	14.24	31	62	14.19
Yien et al. (2011)	16.94	2.38	33	15.09	3.39	33	66	1.85
Chu & Hung (2015)	56	23.54	30	71.03	23.04	29	59	-15.03
Su & Cheng (2015)	82.94	10	34	75.59	9.595	34	68	7.35
Chen & Hwang (2017)	86.78	9.15	27	82.35	12.38	26	53	4.43
Fan et al. (2015)	88	9	23	76	12	23	46	12
Furió et al. (2015)	4.89	1.45	19	4.74	0.93	19	38	0.15
Chen et al. (2016)	18.51	2.71	43	16.63	4	77	120	1.88
	41.12		Σ= 426	37.07		Σ= 456	Σ= 954	

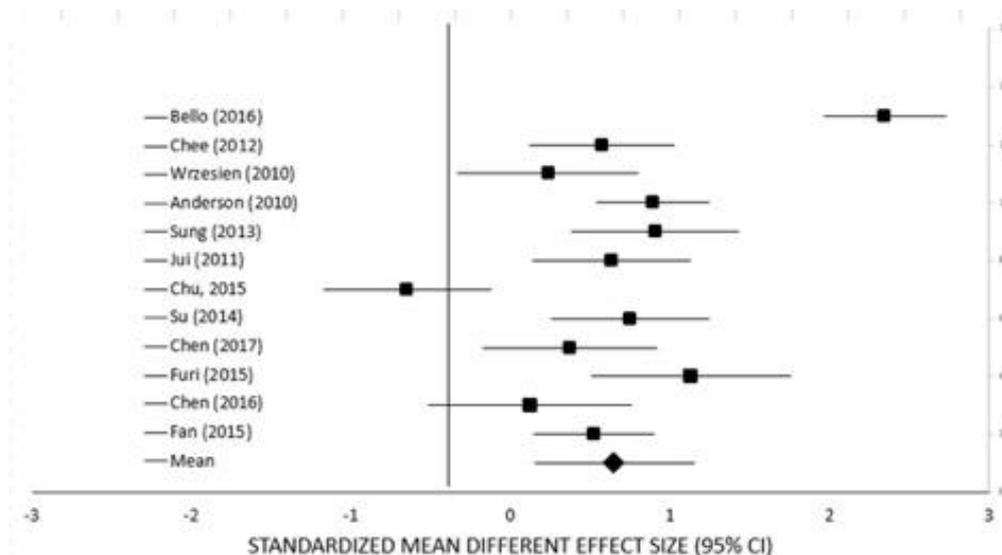


Figure 5. The forest plot's comparison of the SMD effect size of reviewed studies

Table 3. Effect sizes, confidence intervals, and relative weights of reviewed studies

Name of the Study	Hedges's g	Lower Limit	Upper Limit	p-Value	Relative Weight (W)
Bello et al. (2016)	2.338	1.959	2.716	0.00000	8.738
Chee & Tan (2012)	0.571	0.123	1.020	0.01255	8.502
Wrzesien & Raya (2010)	0.233	-0.325	0.792	0.41332	8.089
Anderson & Barnett (2013)	0.886	0.536	1.237	0.00000	8.821
Sung & Hwang (2013)	0.898	0.381	1.414	0.00065	8.253
Yien et al. (2011)	0.624	0.136	1.113	0.01227	8.358
Chu & Hung (2015)	-0.637	-1.153	-0.120	0.01571	8.252
Su & Cheng (2015)	0.741	0.255	1.228	0.00279	8.367
Chen & Hwang (2017)	0.402	-0.134	0.938	0.14151	8.177
Fan et al. (2015)	1.112	0.500	1.724	0.00036	7.873
Furió et al. (2015)	0.121	-0.503	0.744	0.70453	7.826
Chen et al. (2016)	0.520	0.143	0.896	0.00682	8.743
Randon effect model	0.661	0.232	1.090	0.00253	

Table 4. Mean effect size of GBSL based on school level

Moderator	Number of studies	% of study	d	N
Elementary	7	58.33%	1.08	394
Secondary	5	41.67%	0.34	560

Do Moderator Categories Including School Level of Participants (Elementary and Secondary School Context) and Year of Publication Have Any Correlation with GBSL Effect Size?

Based on our analysis of moderating variables as the addition to the overall effect size, subsequent analyses of some moderating variables were run by school level and year of journal article's publication, shown in Table 4.

Firstly, we made two comparisons from the school level including elementary and sec-

ondary schools (Table 4). Seven studies are in the context of an elementary school setting with the mean of effect size 1.08. The other five studies tested on secondary school setting with an effect size mean of 0.34. This number shows that the effect size of GBSL on secondary school contexts nearly two and a half times higher than elementary school students sample effect size. Thus, the implementation of GBSL in secondary school tend to have a larger effect size than in elementary school context.

Secondly, we made a comparison of effect size according to the year of publication (Table 5). According to the correlational analysis between the year of publication and effect size, it shows that the variable has a low correlation with the $r = 0.40$ ($r^2 = 0.16$). Figure 6 illustrates a scatter plot that shows the relationship between year of publication (X-axis) and effect size (y-axis). Figure 6 shows that from 2010 the effect size average is 0.23, followed by approximately double to 0.55 in 2011. Five years later, in 2016, the effect size significantly increased again to 2.54.

Analysis for Publication Bias

According to the analysis of Rosenthal's Fail-safe N (Orwin, 1983), among the various methods for assessing bias, Rosenthal's Fail-safe N has the advantage of focusing on the potential impact any unpublished or unidentified studies may have on the current estimated effect size. It provides an estimate for the number of hypothetical missing studies

that must be identified in order to bring the calculated overall effect below the level of researcher-imposed substantive significance (Easterbrook, Gopalan, Berlin, & Matthews, 1991). It assumes that those missing studies have negligible effects. Based on the analysis, 307 more studies are needed to make p-value to be alpha (Z for alpha = 1.959). The other method to analyze publication bias is using the Funnel Plot, which has two diagonal lines that represent the 95% confidence interval, and a vertical central line. The x-axis represents the study sample size, and the y-axis represents the effect size. Figure 7 illustrates the Funnel plot of Standard Error (SE) by Hedges' g effect size.

According to Figure 7, the nine studies fall around the two horizontal lines or a confidence interval of 95%. However, three studies fall outside the funnel plot, indicating that these studies were not as significant as the other nine studies.

Table 5. Mean effect size of GBSL based on year of publication

Year of Publication	Number of Studies	% of study	d	N
2010	1	8.33%	0.23	48
2011	1	8.33%	0.55	66
2012	1	8.33%	0.75	78
2013	2	16.67%	0.892	198
2014	1	8.33%	0.77	68
2015	3	25.00%	0.51	143
2016	2	16.67%	2.54	300
2017	1	8.33%	0.36	53

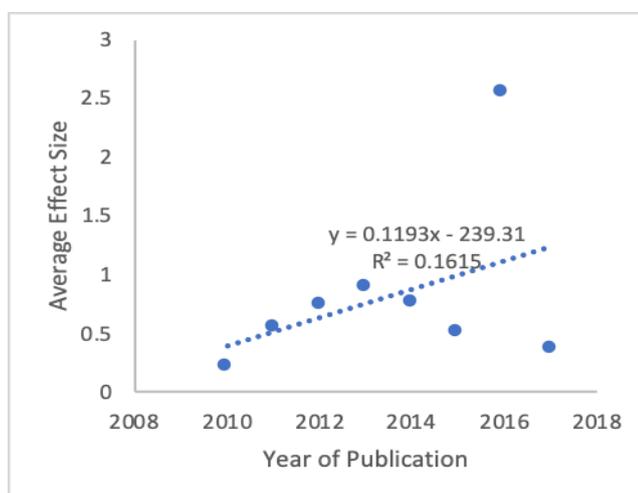


Figure 6. Scatter plot of the relationship between the year of publication and average effect size

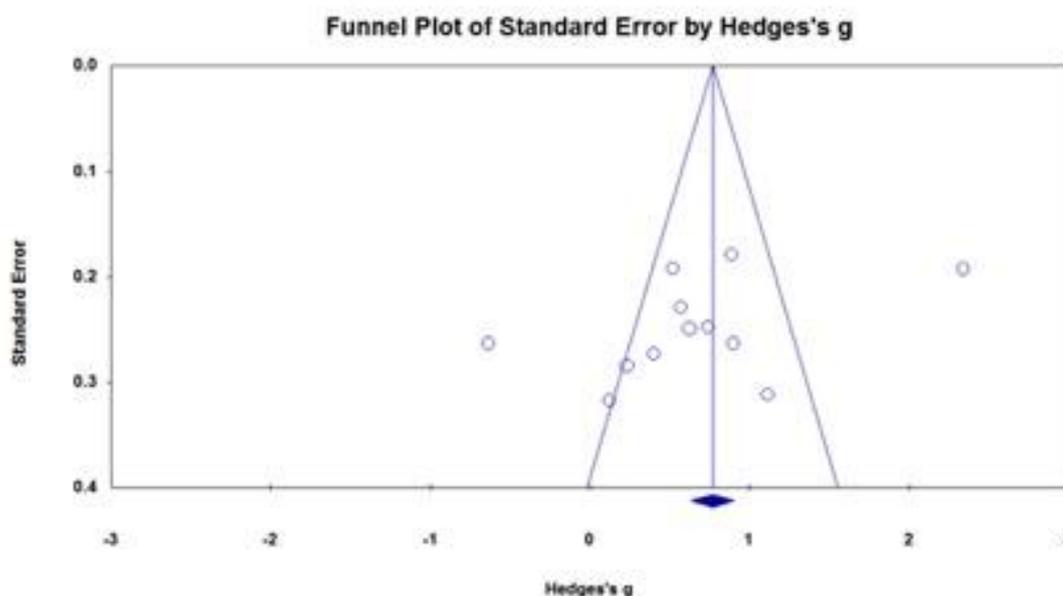


Figure 7. Funnel plot of standard error (SE) by hedges's g effect size of reviewed studies

The Performance of the Result of This Study with Similar Research

The performance of this study aligns with similar studies of literature reviews using meta-analysis on gamification across various context, such as mathematics, language, and also physical education over a decade, which has consistently found that game-based learning outperforms traditional-based learning (Divjak & Tomić, 2011; Vogel et al., 2006; Young et al., 2012). However, some notable differences regarding the statistical analysis are revealed. First, the fail-safe number (Nfs) that we found in this research, that is 307 studies, is much lower than the previous meta-analysis. The fail-safe number is only approximately a fifth than the findings of Vogel et al. (2006) with Nfs 1465. Second, the number of studies in this meta-analysis is only twelve, which is lower than similar research in this field, such as Divjak and Tomić (2011) with 32 studies, and Young et al. (2012) with more than 300 articles. In addition, the findings of this research support the findings of Li and Tsai (2013) regarding the potential of GBSL to promote students' learning. Li and Tsai (2013) believe that GBSL can promote students' engagement. Therefore, students' engagement and motivation might lead to an improvement in students' learning outcomes in Science.

Conclusion and Recommendation

Conclusion

Based on the result and discussion, some conclusions can be drawn. First, based on the investigated studies conducted from 2010 to 2017, the use of GBSL is statistically significant to improve students' learning outcomes in elementary and secondary school. The learning outcome of the experimental group of the overall studies is higher than the control group, which is 41.12 against 37.07 respectively. The mean of Hedges' g random effect size of the reviewed studies is 0.667, which can be classified into a medium effect size. Second, moderator categories or variation of school level of the study have any correlation on digital game effectiveness on which the implementation of GBSL in secondary school have a greater effect size than in elementary school context. Also, the year of publication and effect size has a low positive correlation with $r = 0.40$.

Recommendation

The result of this study has implications for future studies. Experimental research of GBSL in Science education across various contexts is still needed. It is supported by the result of detection publication bias which showed that at least 237 studies in this area of

research are needed that would bring p-value to be alpha. This research is complex, but the description of the process and result has been presented. Furthermore, we use Comprehensive Meta-Analysis 2.0 as trusted software for quantitative meta-analysis.

However, our study has some limitations. The study only includes a small amount of research. It might be caused by the topic used is too specific where it only includes the effect of GBSL in a subject (Science) and the outcomes only specifically focus on cognitive aspects. There are many potential studies in GBSL in Science education and in the timeframe (2010-2017), but they were not included in this study because they were not eligible in the screening process with the seven inclusion and exclusion criteria which is determined in the research design. Some researches have no complete data to be extracted, or the topic is not suitable for this research. For example, the research use case study which only has an experimental group does not have a control group (Echeverría et al., 2011; Spires, Rowe, Mott, & Lester, 2011). Other studies are not eligible because they focus on other outcomes such as engagement (Annetta et al., 2009), collaboration and problem-solving (Sánchez & Olivares, 2011), and developing serious games (Khalili, Sheridan, Williams, Clark, & Stegman, 2011; Nilsson & Jakobsson, 2011; Ting, 2010).

Therefore, future studies should not only focus on the cognitive or quantitative outcome but also affective or qualitative outcomes such as students' engagement, motivation, self-efficacy, participation, collaboration, communication, and problem-solving skills. The research to review the qualitative outcome can be conducted with a systematic review, narrative review, or descriptive review (For example, Kim, Munson, & McKay, 2012; Li & Tsai, 2013).

The limited number of research identified might also due to the restricted criteria of the year of publication, sources of databases, context, and moderator categories. First, the included studies were conducted from 2010 to 2017. Therefore, the result of this study does not capture the studies outside this period. Second, the review only includes

some databases, including ERIC, Springer Link, ProQuest, and A+ Education. Future studies can also be conducted by extending the literature to other educational databases such as ISI Web of Sciences or sources like Google Scholar, conference proceedings, and dissertations. There many articles related to GBSL.

Third, regarding context, investigating the effectiveness in different contexts/country and expanded educational level such as preschool could also be explored in future studies. It is because we found that most of the research included in this meta-analysis was conducted within Asia and educational level in the preschool context has not been explored. The last, for moderator categories, our research only focused on the school level of participants and year of publication of the study. Therefore, future research can explore different moderators such as gender (Tsay et al., 2018; Vogel et al., 2006), game genre (individual, peers, or groups), stream type or typical games (Sjöblom, Törhönen, Hamari, & Macey, 2017), learner control, and type of activity (Vogel et al., 2006).

Acknowledgment

This research is funded by LPDP (*Lembaga Pengelola Dana Pendidikan*). We also acknowledge the lecturer of Research Approaches in Education and Research Project in Education in the Faculty of Education, Monash University Australia who gave many suggestions to this article. Lastly, we thank Monash University Library that provides resources for this article.

References

- Adam, S. (2004). A consideration of the nature, role, application, and implications for European education of employing 'learning outcomes' at the local, national, and international levels. *United Kingdom Bologna Seminar*. 1-2 July 2004, Heriot-Watt University (Edinburgh Conference Centre), Edinburgh.
- Anderson, J. L., & Barnett, M. (2013). Learning physics with digital game

- simulations in middle school science. *Journal of Science Education and Technology*, 22(6), 914–926. <https://doi.org/10.1007/s10956-013-9438-8>
- Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M.-T. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education*, 53(1), 74–85. <https://doi.org/10.1016/j.compedu.2008.12.020>
- Bello, S., Ibi, M. B., & Bukar, I. B. (2016). Effect of simulation techniques and lecture method on students' academic performance in mafoni day secondary school Maiduguri, Borno state, Nigeria. *Journal of Education and Practice*, 7(23), 113–117. Retrieved from <https://www.iiste.org/Journals/index.php/JEP/article/view/32584>
- Bennett, S., Maton, K., & Kervin, L. (2008). The 'digital natives' debate: A critical review of the evidence. *British Journal of Educational Technology*, 39(5), 775–786. <https://doi.org/10.1111/j.1467-8535.2007.00793.x>
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis version 2*. Englewood Cliffs, NJ: Biostat.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97–111. <https://doi.org/10.1002/jrsm.12>
- Borg, W. R., & Gall, M. D. (1983). *Educational research: An introduction* (4th ed.). New York, NY: Longman.
- Brockwell, S. E., & Gordon, I. R. (2001). A comparison of statistical methods for meta-analysis. *Statistics in Medicine*, 20(6), 825–840. <https://doi.org/10.1002/sim.650>
- Chee, Y. S., & Tan, K. C. D. (2012). Becoming chemists through game-based inquiry learning: The case of "legends of alkhimia." *Electronic Journal of E-Learning*, 10(2), 185–198. Retrieved from <http://www.ejel.org/issue/download.html?idArticle=188>
- Chen, C.-L. D., Yeh, T.-K., & Chang, C.-Y. (2016). The effects of game-based learning and anticipation of a test on the learning outcomes of 10th grade Geology students. *EURASIA Journal of Mathematics, Science and Technology Education*, 12(5), 1379–1388. <https://doi.org/10.12973/eurasia.2016.1519a>
- Chen, C., & Hwang, G. (2017). Effects of the team competition-based ubiquitous gaming approach on students' interactive patterns, collective efficacy and awareness of collaboration and communication. *Educational Technology & Society*, 20(1), 87–98. Retrieved from <https://www.jstor.org/stable/jeductech.soci.20.1.87>
- Cheng, M.-T., Su, T., Huang, W.-Y., & Chen, J.-H. (2014). An educational game for learning human immunology: What do students learn and how do they perceive? *British Journal of Educational Technology*, 45(5), 820–833. <https://doi.org/10.1111/bjet.12098>
- Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88–113. <https://doi.org/10.1016/j.edurev.2013.01.001>
- Chorney, A. I. (2012). Taking the game out Of gamification. *Dalhousie Journal of Interdisciplinary Management*, 8(1), 1–14. <https://doi.org/10.5931/djim.v8i1.242>
- Chu, H.-C., & Hung, C.-M. (2015). Effects of the digital game-development approach on elementary school students' learning motivation, problem solving, and learning achievement. *International Journal of Distance Education Technologies*, 13(1), 87–102. <https://doi.org/10.4018/ijdet.2015010105>
- Clark, D. B., Nelson, B. C., Chang, H.-Y., Martinez-Garza, M., Slack, K., &

- D'Angelo, C. M. (2011). Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in Taiwan and the United States. *Computers & Education*, 57(3), 2178–2195. <https://doi.org/10.1016/j.compedu.2011.05.007>
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). New York, NY: Routledge.
- Corbett, S. (2010, September 15). Learning by playing video games in the classroom. *The New York Times*. Retrieved from <https://www.nytimes.com/2010/09/19/magazine/19video-t.html>
- Culp, K. M., Martin, W., Clements, M., & Lewis Presser, A. (2015). Testing the impact of a pre-instructional digital game on middle-grade students' understanding of photosynthesis. *Technology, Knowledge and Learning*, 20(1), 5–26. <https://doi.org/10.1007/s10758-014-9233-5>
- Divjak, B., & Tomić, D. (2011). The impact of game-based learning on the achievement of learning goals and motivation for learning mathematics - literature review. *Journal of Information and Organizational Sciences*, 35(1), 15–30. Retrieved from <https://jios.foi.hr/index.php/jios/article/view/182>
- Easterbrook, P. J., Gopalan, R., Berlin, J. A., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet*, 337(8746), 867–872. [https://doi.org/10.1016/0140-6736\(91\)90201-Y](https://doi.org/10.1016/0140-6736(91)90201-Y)
- Echeverría, A., García-Campo, C., Nussbaum, M., Gil, F., Villalta, M., Améstica, M., & Echeverría, S. (2011). A framework for the design and integration of collaborative classroom games. *Computers & Education*, 57(1), 1127–1136. <https://doi.org/10.1016/j.compedu.2010.12.010>
- Fan, K.-K., Xiao, P., & Su, C. (2015). The effects of learning styles and meaningful learning on the learning achievement of gamification health education curriculum. *EURASIA Journal of Mathematics, Science and Technology Education*, 11(5), 1211–1229. <https://doi.org/10.12973/eurasia.2015.1413a>
- Furió, D., Juan, M.-C., Seguí, I., & Vivó, R. (2015). Mobile learning vs. traditional classroom lessons: A comparative study. *Journal of Computer Assisted Learning*, 31(3), 189–201. <https://doi.org/10.1111/jcal.12071>
- Gee, J. P. (2007). *What video games have to teach us about learning and literacy* (Revised an). New York, NY: Palgrave Macmillan.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage Publications.
- Hamari, J., & Keronen, L. (2017). Why do people play games? A meta-analysis. *International Journal of Information Management*, 37(3), 125–141. <https://doi.org/10.1016/j.ijinfomgt.2017.01.006>
- Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3(4), 486–504. <https://doi.org/10.1037/1082-989X.3.4.486>
- Honey, M. A., & Hilton, M. (Eds.). (2011). *Learning science through computer games and simulations*. Washington, DC: National Academy of Sciences.
- Huang, B., Hew, K. F., & Lo, C. K. (2019). Investigating the effects of gamification-enhanced flipped learning on undergraduate students' behavioral and cognitive engagement. *Interactive Learning Environments*, 27(8), 1106–1126. <https://doi.org/10.1080/10494820.2018.1495653>
- Khalili, N., Sheridan, K., Williams, A., Clark, K., & Stegman, M. (2011). Students designing video games about immunology: Insights for science learning. *Computers in the Schools*, 28(3), 228–240. <https://doi.org/10.1080/07380569.2011.594988>
- Kim, H., Munson, M. R., & McKay, M. M. (2012). Engagement in mental health

- treatment among adolescents and young adults: A systematic review. *Child and Adolescent Social Work Journal*, 29(3), 241–266. <https://doi.org/10.1007/s10560-012-0256-2>
- King, W. R., & He, J. (2005). Understanding the role and methods of meta-analysis in IS research. *Communications of the Association for Information Systems*, 16, 665–686. <https://doi.org/10.17705/1CAIS.01632>
- Li, M.-C., & Tsai, C.-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>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (Vol. 49). Thousand Oaks, CA: Sage Publications.
- Mayo, M. J. (2007). Games for science and engineering education. *Communications of the ACM*, 50(7), 30–35. <https://doi.org/10.1145/1272516.1272536>
- McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. New York, NY: Penguin.
- Mertens, D. M. (2015). *Research and evaluation in education and psychology* (4th ed.). London: SAGE Publication.
- Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews*, 82(4), 591–605. <https://doi.org/10.1111/j.1469-185X.2007.00027.x>
- Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6), 1417–1426. <https://doi.org/10.2307/1911408>
- Nilsson, E. M., & Jakobsson, A. (2011). Simulated sustainable societies: Students' reflections on creating future cities in computer games. *Journal of Science Education and Technology*, 20(1), 33–50. <https://doi.org/10.1007/s10956-010-9232-9>
- Okeke, G. N. (2016). *The impact of digital games on high school students' learning outcomes in mathematics education: A meta-analytic investigation*. Doctoral thesis, University of North Texas, Denton, TX.
- Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, 25(3), 241–286. <https://doi.org/10.1006/ceps.2000.1040>
- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, 78(4), 679–703. <https://doi.org/10.1037/0021-9010.78.4.679>
- Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, 8(2), 157–159. <https://doi.org/10.2307/1164923>
- Papastergiou, M. (2009). Digital game-based learning in high school Computer Science education: Impact on educational effectiveness and student motivation. *Computers & Education*, 52(1), 1–12. <https://doi.org/10.1016/j.compedu.2008.06.004>
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *JAMA*, 295(6), 676. <https://doi.org/10.1001/jama.295.6.676>
- Prensky, M. (2001). Digital natives, digital immigrants Part 1. *On the Horizon*, 9(5), 1–6. <https://doi.org/10.1108/10748120110424816>
- Quandt, T., Van Looy, J., Vogelgesang, J., Elson, M., Ivory, J. D., Consalvo, M., & Mäyrä, F. (2015). Digital games research: A survey study on an emerging field and its prevalent debates. *Journal of Communication*, 65(6), 975–996. <https://doi.org/10.1111/jcom.12182>
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2006). *Publication bias in meta-*

- analysis: *Prevention, assessment and adjustments*. Hoboken, NJ: John Wiley & Sons.
- Sacks, H. S., Berrier, J., Reitman, D., Ancona-Berk, V. A., & Chalmers, T. C. (1998). Meta-Analyses and Large Randomized, Controlled Trials. *New England Journal of Medicine*, 338(1), 59–62. <https://doi.org/10.1056/NEJM199801013380112>
- Sánchez, J., & Olivares, R. (2011). Problem solving and collaboration using mobile serious games. *Computers & Education*, 57(3), 1943–1952. <https://doi.org/10.1016/j.compedu.2011.04.012>
- Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings*. London: SAGE Publications.
- Sjöblom, M., Törhönen, M., Hamari, J., & Macey, J. (2017). Content structure is king: An empirical study on gratifications, game genres and content type on Twitch. *Computers in Human Behavior*, 73, 161–171. <https://doi.org/10.1016/j.chb.2017.03.036>
- Spires, H. A., Rowe, J. P., Mott, B. W., & Lester, J. C. (2011). Problem solving and game-based learning: Effects of middle grade students' hypothesis testing strategies on learning outcomes. *Journal of Educational Computing Research*, 44(4), 453–472. <https://doi.org/10.2190/EC.44.4.e>
- Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., ... Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 343(jul22 1), d4002–d4002. <https://doi.org/10.1136/bmj.d4002>
- Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis. *Journal of Clinical Epidemiology*, 54(10), 1046–1055. [https://doi.org/10.1016/S0895-4356\(01\)00377-8](https://doi.org/10.1016/S0895-4356(01)00377-8)
- Su, C.-H., & Cheng, C.-H. (2015). A mobile gamification learning system for improving the learning motivation and achievements. *Journal of Computer Assisted Learning*, 31(3), 268–286. <https://doi.org/10.1111/jcal.12088>
- Sung, H.-Y., & Hwang, G.-J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education*, 63, 43–51. <https://doi.org/10.1016/j.compedu.2012.11.019>
- Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., & Song, F. (2000). *Methods for meta-analysis in medical research*. Chichester: John Wiley & Sons.
- Ting, Y. L. (2010). Using mainstream game to teach technology through an interest framework. *Journal of Educational Technology & Society*, 13(2), 141–152. Retrieved from <https://scholar.lib.ntnu.edu.tw/en/publications/using-mainstream-game-to-teach-technology-through-an-interest-fra>
- Tsay, C. H.-H., Kofinas, A., & Luo, J. (2018). Enhancing student learning experience with technology-mediated gamification: An empirical study. *Computers & Education*, 121, 1–17. <https://doi.org/10.1016/j.compedu.2018.01.009>
- van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. *EDUCAUSE Review*, 41(2), 16–30. Retrieved from <https://er.educause.edu/articles/2006/1/digital-gamebased-learning-its-not-just-the-digital-natives-who-are-restless>
- Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, 34(3), 229–243. <https://doi.org/10.2190/FLHV-K4WA-WPVQ-H0YM>
- Wrzesien, M., & Raya, M. A. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project.

Computers & Education, 55(1), 178–187.
<https://doi.org/10.1016/j.compedu.2010.01.003>

Yien, J.-M., Hung, C.-M., Hwang, G.-J., & Lin, Y.-C. (2011). A game-based learning approach to improving students' learning achievements in a nutrition course. *TOJET: The Turkish Online Journal of Educational Technology*,

10(2). Retrieved from <http://www.tojet.net/articles/v10i2/1021.pdf>

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., ... Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research*, 82(1), 61–89. <https://doi.org/10.3102/0034654312436980>