



## THE EFFECTIVENESS OF GAME-BASED SCIENCE LEARNING (GBSL) TO IMPROVE STUDENTS' LEARNING OUTCOME: A META-ANALYSIS OF CURRENT RESEARCH FROM 2010 TO 2017

Heru Setiawan<sup>1,✉</sup>

<sup>1</sup> Biology Teacher in Senior High of Sekolah Global Mandiri, Jakarta

### Article Info

#### Article History:

Received April 2019

Accepted August 2019

Published December 2019

#### Keywords:

Game-based science learning,  
learning outcomes, meta-  
analysis

### Abstract

This study identifies the effectiveness of the 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. Method: 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. Findings: (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 low positive correlation with a coefficient of correlation 0.40.

© 2019 Universitas Negeri Semarang

p-ISSN 2252-6617

e-ISSN 2252-6232

✉Correspondence author:

**Heru Setiawan**

Senior High of Sekolah Global Mandiri

Jl. Raya Cakung Cilincing Km.05 Jakarta Garden City, Cakung Jakarta Timur, 13910, Jakarta, Indonesia

Email: [heru.setiawan@teacher.globalmandiri.sch.id](mailto:heru.setiawan@teacher.globalmandiri.sch.id)

## INTRODUCTION

The young generation who was born in the 21st century is digital native or the Net generation (Bennett, Maton & Kervin, 2008). The millennial in this era also can be called game generation (Prensky, 2001). The trend of digital games' use has been increasing in this era (McGonigal, 2011; Corbett, 2012). Millions of people have been immersed in playing digital games either for entertainment or education (Huang, Hew, & Lo, 2018). Gee (2007) reported in his study that approximately 90% of students' mobile phone connect to digital games. Besides, many teachers use digital games as a medium of instruction in their classroom for engaging students during teaching and learning processes, or it is commonly called Digital game-based learning (DGBL) (Van, 2006; Papastergiou, 2009). Students also obtain feedbacks such as improvement and win condition after completing the goals (Okeke, 2016, p. 1). The DGBL that specifically focus on Science can be called Game Based Science Learning (GBSL).

Since 2006, the number of research investigating the effect of Digital game 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, 2015). They argue that science is challenging for some students because of abstract concepts and invisible objects. Also, some research illustrated that rote memorization and decontextualized learning have potential drawbacks in 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 (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, Hew, & Lo, 2018; 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 between 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 outcome 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 treatment group (Nakagawa & Cuthill, 2007). Meanwhile, the small sample size of the research cannot be used to generalize the effect of GBSL. To solve this issue, it needs further investigation of the effectiveness of GBSL in students' achievement in sciences with meta-analysis study to develop a better estimate of effect magnitude (King & He, 2005). Meta-analysis is as the process of converting the effects of several similar research into a 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, Smith & McGaw, 1981; Hattie, 2009). 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. 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. While, learner control, type of activity, and realism do not appear to be influential. In science context, Li & 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 quantitative approach. Li & Tsai (2013) who focused their research on the qualitative method suggested that quantitative content analysis of GBSL effectiveness such as students' learning outcome in Science education should be conducted in future investigations. This is because digital games that can promote students' engagement (Tsay, Kofinas, & Luo, 2018; Annetta, Minogue, Holmes & Cheng, 2009) might also enhance students' learning outcomes (Prensky, 2002). The 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:

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?

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. Meta-analysis of GBSL has been conducted by several researchers within a broader context such as mathematics, language and other subjects (Divjak & Tomic, 2011; Young et al., 2012), but lack of

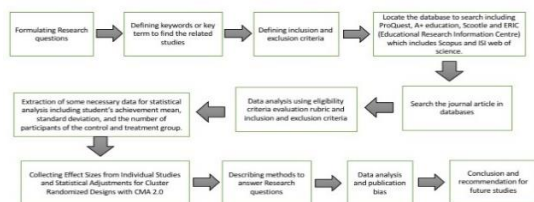
research conducted in science education. Second, the consistency of the result of similar studies for several years will be investigated. Therefore, consistency and inconsistency of finding of similar research will be found, and bias of one or more studies in this field could be detected (Borg & Gall, 1983). Third, 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, Smith & McGaw, 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 will create "greater statistical power and more precise confidence intervals" (Levine et al., 2008, p. 202). 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 (Mark, Lipsey & Wilson, 2001). Also, through the substantial number of participant 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, we can investigate different moderator variables. Vogel et al., (2006) argue that analyzing moderator variables would give a clearer overview or more complex picture of reviewed studies.

## METHODS

Method used should be accompanied by references, relevant modification should be explained. Procedure and data analysis technique should be emphasized to literature review article.

### Diagram of the Research Process

Figure 1 shows a diagram of the research process including an identification of reviewed studies.



**Figure 1.** The selection process to identify studies for the inclusion and exclusion process

### Search 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 that will be included in this study including publication year, unit, participant, game intervention, research design, participant, outcome type, and language. These detail of inclusions and exclusions are explained as follows.

1. Publication year: All of the articles are peer-reviewed journal article which was 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 subject 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 subject 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 Journal articles included in this meta-analysis must use experimental and control groups or game versus non-game condition. 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 disability 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 where 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 includes 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 search 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 mean, standard deviation, the number of participants of the control and treatment group.

### Data Analysis Method

Microsoft Excell and Comprehensive Meta-Analysis (CMA 2.0) were used for statistical analysis after the quantitative data were extracted. The former, the demographic characteristics of the reviewed studies are analyzed with descriptive statistics using Microsoft Excell which present data such as mean, percentage, and frequencies. The data will also be presented with visual techniques such as a column, bar chart, and histogram. The latter, CMA 2.0 was also used. Several researchers verified the accuracy of the analysis method (Ones, Viswesvaran, & Schmidt, 1993; Hülshager, Anderson & Salgado, 2009). CMA 2.0 is used to analyze hedges'  $g$  effect size, the lower limit (LL), the upper limit (UL),  $p$ -Value, the Relative weight of all studies (Borenstein, Hedges, Higgins, & Rothstein, 2005). 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 et al., 2000). Two kinds of effect model in a meta-analysis are fixed effect model and random effect model (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 conclusion (Nickell, 1981). However, in this study, we use the random effect model because all twelve studies used in this research are drawn from different population, such as different populations in different countries. The similar condition of research is conducted by Sacks et al. (1987). Moreover, the studies report varies the effect size (ES). In the random-effects model, the true effect size might differ from study to 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 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 will 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 control group. The second research question is answered by investigating the effect of moderator categories including year and school level, to the GBSL effectiveness, we use descriptive analysis through comparing the mean of effect size in each category. We compare the average effect size in each school level (elementary and secondary school) to determine which school level more effective of 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 will be presented as scatterplot to illustrate the relationship between two variables (Cohen, Manion & Morrison, 2007, p. 507). It will 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 are 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

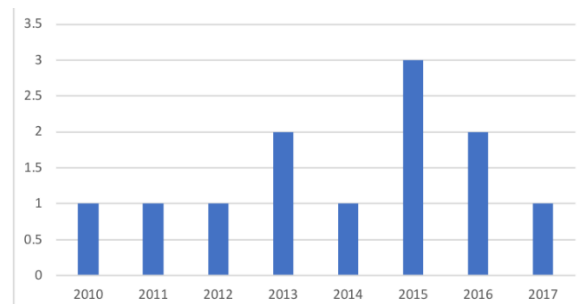
Detection of publication bias of reviewed studies is crucial in meta-analysis study (Rothstein, Sutton & Borenstein, 2006). Publication bias is the tendency of researchers to screen article for publication based on the statistical significance of effects than the quality of the study (Rothstein, Sutton, & Borenstein,

2006, p. 296). Several pieces of evidence show that some research that has higher effect size is more likely to be published (Peters, Sutton, Jones, Abrams & Rushton, 2006). Consequently, it will affect the review process. So, it is possible that the meta-analysis is overestimated effect size because it uses a biased sample or target of the population. So, 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 (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 (Becker, 2005). Orwin (1983) suggested that Rosenthal's Fail-safe N compute the number of studies that should be incorporated in the analysis.

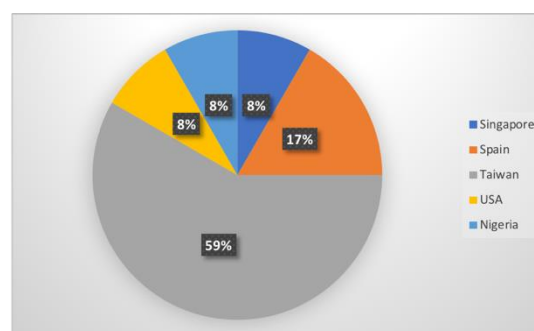
## RESULTS AND DISCUSSION

### Overview of the Reviewed Studies

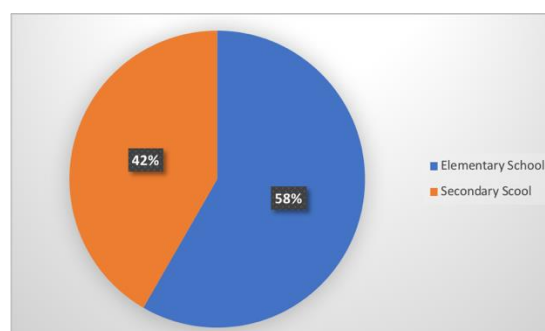
The publication years range from 2010 to 2017. The purpose is to know the development of research in this area in the last eight years. The highest number of publication is in 2015 with three publications (Figure 2). Then, the presence of international studies is reflected in the sample. However, 50% of the studies included were conducted within Asia continent especially in Taiwan, while the others were conducted internationally. There are two countries including Taiwan and Singapore from Asia. Within this international group, Spain is well represented by two studies. While the other research is from the U.S and Nigeria, Africa (Figure 3). 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 4). 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 5).



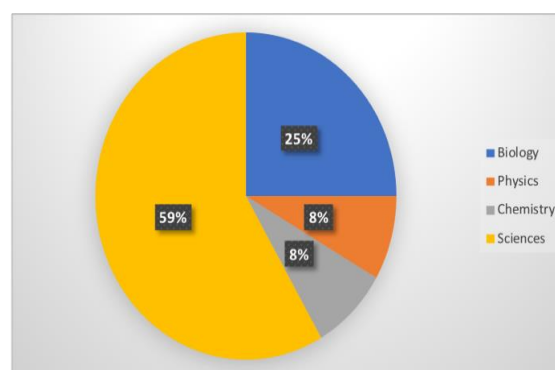
**Figure 2.** The number of reviewed studies by year of publication



**Figure 3.** The reviewed studies by country



**Figure 4.** The reviewed studies based on the education level of participants



**Figure 5.** The reviewed studies according to Science domain

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

**Table 1.** The Background information of reviewed articles

Authors	Year of publication	Country	Game name	School level	Science domain
Bello, Ibi & Bukar	2016	Nigeria	N/a	Secondary	Sciences
Chee & Tan	2012	Singapore	Alkhimia	Secondary	Chemistry
Wrzesien & Alcañiz Raya	2010	Spain	Supercharged	Elementary	Sciences
Anderson & Barnett	2013	USA	Supercharged	Secondary	Physics
Sung & Hwang	2013	Taiwan	Alien Invasion	Elementary	Sciences
Jui-Mei, Chun-Ming, Hwang & Yueh-Chiao	2011	Taiwan	Nutrition Supplement Battle	Elementary	Biology
Chu & Hung	2015	Taiwan	Kodu	Elementary	Sciences
Su & Cheng	2014	Taiwan	Find Insect	Elementary	Sciences
Chen & Hwang	2017	Taiwan	Alien Invasion	Elementary	Sciences
Fan, Xiao, & Su	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 Game-Based Science Learning (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 will be 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 treatment group and control group. The number of participant in the twelve studies are 954 students. Most of the studies have an equal number of participant 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 participant 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.

**Table 2.** Mean, Standard Deviation and Sample Size of the twelve Studies on Digital Games Versus Control Method

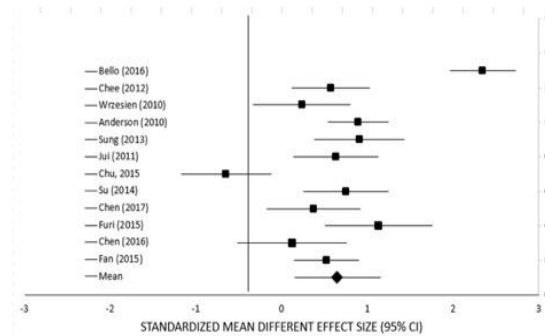
Main Author (year)	Experiment class			Control Class			N total	Mean difference
	Mean	SD	N	Mean	SD	N		
Bello (2016)	66.23	7.07	90	46.6	9.48	90	180	19.63
Chee (2012)	3.28	2.61	40	2	1.71	38	78	1.28
Wrzesien (2010)	6.33	2.2	24	5.88	1.54	24	48	0.45
Anderson (2013)	6.3	1.2	32	5.9	1.27	32	136	0.4
Sung (2013)	57.26	16.87	31	43.07	14.24	31	62	14.19
Jui (2011)	16.94	2.38	33	15.09	3.39	33	66	1.85
Chu, 2015	56	23.54	30	71.03	23.04	29	59	-15.03
Su (2014)	82.94	10	34	75.59	9.595	34	68	7.35
Chen (2017)	86.78	9.15	27	82.35	12.38	26	53	4.43
Fan (2015)	88	9	23	76	12	23	46	12
Furió (2015)	4.89	1.45	19	4.74	0.93	19	38	0.15
Chen (2016)	18.51	2.71	43	16.63	4	77	120	1.88
	11.12		426	37.07		Σ= 456	Σ= 954	



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 (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 different 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)

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 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. This 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, Ibi & Bukar (2016) of 2.338. In contrast, the study contributing the smallest overall influence is Chu & 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 6.



**Figure 6.** The comparison of the SMD effect size of reviewed studies

Table 3. Effect Sizes, Confidence Intervals and Relative Weights of reviewed Studies

Study Name	Hedges's <i>g</i>	Lower limit	Upper limit	p-Value	Relative weight (W)
Bello (2016)	2.338	1.959	2.716	0.000000	8.738
Chee (2012).	0.571	0.123	1.020	0.012552	8.502
Wrzesien (2010)	0.233	-0.325	0.792	0.413325	8.089
Anderson (2013)	0.886	0.536	1.237	0.000001	8.821
Sung (2013)	0.898	0.381	1.414	0.000657	8.253
Jui (2011)	0.624	0.136	1.113	0.012277	8.358
Chu, 2015	-0.637	-1.153	-0.120	0.015714	8.252
Su (2014)	0.741	0.255	1.228	0.002799	8.367
Chen (2017)	0.402	-0.134	0.938	0.141517	8.177
Fan (2015)	1.112	0.500	1.724	0.000368	7.873
Furi (2015)	0.121	-0.503	0.744	0.704539	7.826
Chen (2016)	0.520	0.143	0.896	0.006825	8.743
Randon effect model	0.661	0.232	1.090	0.002538	

**Do moderator categories including school level of participants (elementary and secondary school context) and year of publication has 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 which is presented in table 4.

**Table 4.** Mean Effect Size of GBSL based on Moderator Categories

Moderator	Number of studies	Percentage of study	Mean d	df (N-1)
<b>Reviewed studies by School level of participants</b>				
Elementary	7	58.33%	1.08	393
Secondary	5	41.67%	0.34	559
<b>Reviewed studies by year of publication</b>				
2010	1	8.33%	0.23	47
2011	1	8.33%	0.55	65
2012	1	8.33%	0.75	77
2013	2	16.67%	0.892	197
2014	1	8.33%	0.77	67
2015	3	25.00%	0.51	142
2016	2	16.67%	2.54	299
2017	1	8.33%	0.36	52

Firstly, we made two comparisons by school level including elementary and secondary schools. Seven studies are in the context of an elementary school setting with the mean of effect size 1.08. Other five studies tested on secondary school setting with an effect size mean of 0.34. This number shows that effect size of GBSL on secondary school context nearly two and a half time higher than elementary school students sample effect size. So, the implementation of GBSL in secondary school tend to have a larger effect size than in elementary school context. Secondly, according to the correlational analysis between the year of publication and effect size shows that the variable has a low correlation with the  $r = 0.40$  ( $r^2 = 0.16$ ). Figure 7 illustrates a scatter plot that shows the relationship between year of publication (X-axis) and effect size (y-axis). Figure 7 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.

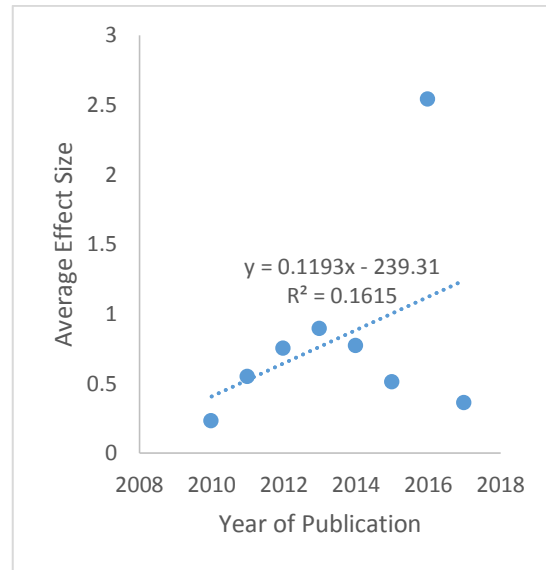
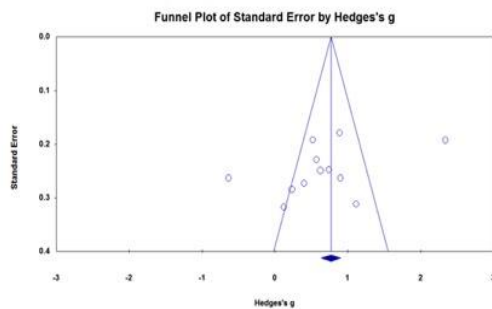


Figure 7. Scatter Plot the relationship between year of publication and average effect size

### 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 Failsafe 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 to bring the calculated overall effect below a level of researcher-imposed substantive significance (Easterbrook, Berlin, Gopalan, & Matthews, 1991). This 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 analysis publication bias is 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 8 illustrates the Funnel plot of Standard error (SE) by hedges' g effect size.



**Figure 8.** Funnel plot of Standard error (SE) by hedges's g effect size of reviewed studies

According to figure 8, 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.

#### **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 physical education, over a decade which have consistently found that games-based learning outperform traditional-based learning (Divjak & Tomić, 2011; Vogel et al., 2006; Young et al., 2012). However, there are some notable differences regarding the statistical analysis. First, the fail-safe number ( $N_f$ s) that we found in this research, 307 studies, is much lower than the previous meta-analysis. The fail-safe number is only approximately a fifth than the findings of Vogel (2006) with  $N_f$ s 1465. Second, the number of studies in this meta-analysis is only twelve which lower than similar research in this field such as Divjak (2011) with 32 studies and Young (2012) with more than 300 articles. Also, the findings of this research support the finding of Li & Tsai (2013) regarding the potential of GBSL to promote students' learning. Li & Tsai (2013) found that GBSL can promote students' engagement. Therefore, students' engagement and motivation might lead to an improvement in students' learning outcomes in Science

## **CONCLUSION**

Based on the result and discussion, we are going to conclude and recommend some possible topic that can be explored in the area of research as the implications of these studies. 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 medium effect size. Second, moderator categories or variation of school level of the study have any correlation on digital game effectiveness where 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 low positive correlation with  $r=0.40$ .

## **RECOMMENDATION FOR FUTURE STUDIES**

The result of this study has implications for future studies. Experimental research of GBSL in Science education across various of context is still needed. This 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 number of research. This 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 of 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 research has 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 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, Minogue, Holmes & Cheng, 2009), collaboration and problem-solving (Sánchez & Olivares, 2011), and developing serious games (Nilsson & Jakobsson, 2011; Khalili et al., 2011; Ting, 2010). So, 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, Li & Tsai, 2013; Kim, Munson & McKay, 2012).

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. So, 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 proceeding, and dissertation. There many articles related to GBSL. Third, regarding context, investigating the effectiveness in different context/country and expanded educational level such as preschool could also be explored in future studies. This 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 school level of participants and

year of publication of the study. Therefore, future research can explore different moderator such as gender (Tsay, Kofinas, & Luo, 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

We Acknowledge Dr Shane Phillipson from Monash University Australia for the help of proofreading 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.
- 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.
- \*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.
- 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.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive*

- meta-analysis version 2. *Englewood, NJ: Biostat*, 104.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed- effect and random- effects models for meta-analysis. *Research synthesis methods*, 1(2), 97-111.
- Borg, W. R., & Gall, M. D. (1983). Educational research an introduction. New York and London.
- Brockwell, S. E., & Gordon, I. R. (2001). A comparison of statistical methods for meta- analysis. *Statistics in medicine*, 20(6), 825-840.
- \*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.
- \*Chen, C. D., Yeh, T., & Chang, C. (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 & Technology Education*, 12(5), 1379-1388.
- \*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.
- 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.
- Cheung, A. C., & 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.
- Chorney, A. I. (2012). Taking the Game Out of Gamification. *Dalhousie Journal of Interdisciplinary Management*, 8(1).
- \*Chu, H., & Hung, C. (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.
- Clark, D. B., Nelson, B. C., Chang, H., 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.
- Cohen, L. Manion, L. and Morrison, K. (2007). *Research methods in education 6th edition*. London: Routledge.
- Corbett, S. (2012). Learning by playing video games in the classroom. *The New York Times*.
- 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.
- 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.
- Easterbrook, P. J., Gopalan, R., Berlin, J. A., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet*, 337(8746), 867-872.
- 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.

- \*Fan, 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 & Technology Education*, 11(5), 1211-1229.
- \*Furió, D., Juan, M. -, Seguí, I., & Vivó, R. (2015). Mobile learning vs. traditional classroom lessons: A comparative study. *Journal of Computer Assisted Learning*, 31(3), 189-201.
- Gee, J. P. (2007). What video games have to teach us about learning and literacy. (Revised and updated edition ed.). New York, N.Y.: Palgrave Macmillan.
- Glass, G. V., Smith, M. L., & McGaw, B. (1981). *Meta-analysis in social research*. Sage Publications, Incorporated.
- Hamari, J., & Keronen, L. (2017). Why do people play games? A Meta-Analysis. *International Journal of Information Management*, 37(3), 125-141.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, 3(4), 486.
- Honey, M. A, Hilton, M. (eds). (2011). Learning science through computer games and simulations. Washington: National Academy of Sciences.
- Huang, B., Hew, K. F., & Lo, C. K. (2018). Investigating the effects of gamification-enhanced flipped learning on undergraduate students' behavioral and cognitive engagement. *Interactive Learning Environments*, 1-21.
- \*Jui-Mei YIEN, Chun-Ming, H., HWANG, G., & Yueh-Chiao, L. I. N. (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).
- Ke, F. (2016). Designing and integrating purposeful learning in game play: A systematic review. *Educational Technology Research and Development*, 64(2), 219-244.
- 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.
- 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.
- 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(1), 32.
- Levine, T. R., Bresnahan, M. J., Park, H. S., Lapinski, M. K., Wittenbaum, G. M., & Shearman, S. M. et al. (2003). Self-construal scales lack validity. *Human Communication Research*, 29, 210-252.
- 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.
- Mark W. Lipsey, & 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.
- McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. 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.

- Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, 1417-1426.
- 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.
- Okeke G N. (2016). The Impact of Digital Games on High School Students' Learning outcomes in Mathematics Education: A Meta-Analytic Investigation. Dissertation: University of North Texas.
- Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary educational psychology*, 25(3), 241-286.
- 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.
- Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of educational statistics*, 8(2), 157-159.
- 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.
- 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), 676-680.
- Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the horizon*, 9(5), 1-6.
- Quandt, T., van Looy, J., Vogelgesang, J., Elson, M., Ivory, J. D., Consalvo, M., et al. (2015). Digital games research: A survey study on an emerging field and its prevalent debates. *Journal of Communication*, 65(6), 975-996.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons.
- Sacks, H. S., Berrier, J., Reitman, D., Ancona-Berk, V. A., & Chalmers, T. C. (1987). Meta-analyses of randomized controlled trials. *New England Journal of Medicine*, 316(8), 450-455.
- Sánchez, J., & Olivares, R. (2011). Problem solving and collaboration using mobile serious games. *Computers & Education*, 57(3), 1943-1952.
- 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.
- 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.
- Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of clinical epidemiology*, 54(10), 1046-1055.
- Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., ... & Tetzlaff, J. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomized controlled trials. *Bmj*, 343, d4002.
- \*Su, C., & Cheng, C. (2015). A mobile gamification learning system for improving the learning motivation and

achievements. *Journal of Computer Assisted Learning*, 31(3), 268-286.

- \*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.
- Sutton, A. J., Abrams, K. R., Jones, D. R., Jones, D. R., Sheldon, T. A., & Song, F. (2000). Methods for meta-analysis in medical research.
- Ting, Y. L. (2010). Using mainstream game to teach technology through an interest framework. *Journal of Educational Technology & Society*, 13(2).
- 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.
- Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. *EDUCAUSE review*, 41(2), 16.
- 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.
- \*Wrzesien, M., & Alcañiz Raya, M. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the e-junior project. *Computers & Education*, 55, 178-187.
- 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.