

The Effects of Digital Game-based Learning Task in English as a Foreign Language Contexts: A Meta-analysis

Chian-Wen KAO

*Department of Applied Foreign Languages
National Taipei University of Business*

A meta-analysis was conducted to investigate the effects of digital game-based learning (DGBL) in English as a foreign language (EFL) contexts. Twenty-five studies were included and they yielded a medium positive effect size ($d = 0.695$, $p < .05$ under the fixed-effect model; $d = 0.777$, $p < .05$ under the random-effects model), suggesting DGBL to be more effective than traditional instruction such as grammar translation methods or audio-lingual methods in EFL contexts. A coding scheme was developed based on the relationship between task-based language learning and DGBL. Six potential moderating variables in three task characteristics were analyzed to further investigate the DGBL effectiveness. Among these potential moderating variables, the treatment duration and linguistic knowledge in task involvement characteristics were found to be influential in DGBL effectiveness. This meta-analysis suggests that digital games should be effective for the acquisition of procedural knowledge of EFL learners who receive the long treatment duration of DGBL. Future research should explore further distinctions in other potential moderating variables in order to provide a more detailed picture of what sorts of variables are influential and what sorts are not.

Keywords: digital game-based learning; meta-analysis; English as a Foreign Language

Whether computers can facilitate learners' acquisition of foreign languages has been the backbone of the research fields of computer-assisted language instruction (CALI) and computer-assisted language learning (CALL). However, second language acquisition (SLA) research in these fields has appeared unable to show much benefit in using computers over

traditional instruction methods (Christensen, Merrill, & Yanchar, 2007; M. Liu, Moore, Graham, & Lee, 2003; Oberg, 2011; Warschauer & Kern, 2000). Garrett (1987), for example, argued that students do not acquire more second language (L2) knowledge from computer-based instruction than traditional workbook exercises because CALI or CALL programs do not provide learners with enough opportunities to interact with the linguistic environment by using linguistic knowledge in real language use. Instead, the programs are simply designed to evaluate whether students can acquire discrete items of linguistic knowledge such as reciting vocabulary or memorizing grammar rules.

English as a foreign language (EFL) learners seldom have exposure to the target language in their daily life. The traditional instruction such as grammar translation methods or audio-lingual methods seems to become a convenient way of developing their target language in classrooms. English education has long been considered as a subject test in the EFL context and teachers' instruction tends to be exam-oriented and drill-driven. However, it has been claimed that learning occurs more when language learners interact to negotiate meaning (Long, 1996). Digital games are now acknowledged as having potential to engage learners and to encourage interaction in the target language (Reinders, 2012). In addition, digital games have been shown to increase intrinsic motivation to learn due to the elements of competition and winning (T. Y. Liu & Chu, 2010). Specifically, digital game-based learning (DGBL) has been incorporated into several content areas to increase learning motivation. Taken as a whole, DGBL has been found more effective than traditional instruction (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Liao, Chang, & Chang, 2010; Wouters & van Oostendorp, 2013); however, when compared individually, DGBL studies yield conflicting results. This has led to an argument that individual content areas must be taken into consideration when assessing the benefits of technology for teaching and learning (Ferdig, 2006). The effectiveness of DGBL might depend on what learning subjects or tasks learners are engaged in. This meta-analysis is therefore conducted to synthesize research which investigates the effects of DGBL in the EFL context. In addition, this meta-analysis begins from the assumption that it is misleading to simply ask the broad question: Do digital games lead to the improvement in EFL learners' language performance? This is roughly like asking whether medication helps alleviate illness. It would be completely surprising to find any evidence for the effectiveness of medication in a study which conflated various medications into one undifferentiated category administered to a conflated group of patients undifferentiated in their types of ailments and then compared its effectiveness to a "control" group that received no medication.

The purpose of the meta-analysis is therefore to analyze several potential moderating variables to investigate which variable is influential in determining the effects of DGBL. To provide pedagogical implications for future researchers or teachers, it is noteworthy that the potential moderating variables investigated in this meta-analysis are from more pedagogical viewpoints. The following research questions are therefore proposed:

1. Do digital games effectively lead EFL learners to the improvement of English learning?
2. Are certain moderating variables more influential than others in determining the effects of DGBL?

Previous Meta-analyses on CALI/CALL Studies

Meta-analysis is a means of synthesizing quantitative results across several empirical studies that address the same or closely related research issues. When the data are produced from experimental or quasi-experimental studies, reviewing an accumulation of such data by vote-counting is limited. In other words, calculating how many studies showed statistically significant positive effects, negative effects or no significant effects for a particular variable or for an intervention in language education is limited. As Norris and Ortega (2000) pointed out, vote-counting cannot take into account differences across the studies such as variations in a sample size. Meta-analysis takes such differences across studies into account and renders the data from various studies representable on a single common scale, making it possible both to provide a weighted effect size taking into account these differences across studies, and to investigate whether certain potential moderating variables are influential in the effectiveness of certain interventions.

There have been numerous meta-analyses published in the research fields of CALI and CALL (e.g., Abraham, 2008; Chang & Lin, 2013; Chiu, 2013; Grgurović, Chapelle, & Shelley, 2013; Lin, Huang, & Liou, 2013; Perez, Noortgate, & Desmet, 2013; Zhao, 2003). Table 1 shows recent meta-analyses on CALI/CALL from 2005 to 2013 and the issues they have addressed. In addition, the number of studies they synthesized and the number of moderator variables they analyzed are displayed.

As shown in Table 1, there are six meta-analyses that have examined the issue of computer or technology use in L2 learning in recent years. Since Felix (2005) suggested that investigating certain particular variables in a meta-analysis should provide more insights into CALI/CALL research fields, most researchers have conducted meta-analyses to investigate the effects of specific forms of technology use on L2 learning and their

Table 1: Summary of Recent Meta-analyses on CALI/CALL Studies

Researchers	Issues addressed	The number of included studies	The number of variables analyzed
Abraham (2008)	The effects of computer-mediated glosses	11	3
Chang & Lin (2013)	The effects of strategy-oriented Web-based English instruction	31	14
Chiu (2013)	The effects of computer-assisted L2 vocabulary instruction	16	4
Grgurović et al. (2013)	The effects of language teaching supported with computer technology	37	9
Lin et al. (2013)	The effects of text-based synchronous computer-mediated communication	10	5
Perez et al. (2013)	The effects of captioned video	18	2

meta-analyses have shown an encouraging result favoring the use of technology. For example, Abraham (2008) investigated the effects of computer-mediated glosses on L2 reading comprehension and vocabulary acquisition. He found a medium effect size of computer-mediated glosses on reading comprehension but a large effect size on vocabulary acquisition. In addition, Perez et al.'s (2013) meta-analysis examined the effects of captioned video on L2 listening comprehension and vocabulary acquisition. Their findings showed a large effect of captioned video on both listening and vocabulary learning. Additionally, Lin et al. (2013) investigated the effects of text-based synchronous computer-mediated communication (SCMC) on SLA.¹ They found a small positive effect size favoring text-based SCMC. Chang and Lin (2013) examined the effects of strategy-oriented Web-based instruction in EFL contexts. Their findings revealed certain strategies in Web-based environment yielded higher effect sizes, such as predicting, summarizing, self-questioning, and so on.

Recall that the research focus of the current meta-analysis is on the effects of DGBL in EFL contexts. Among previous meta-analyses, only Chiu's (2013) meta-analysis was found to relate to the present meta-analysis in the scope of research covered. Her meta-analysis investigated the effects of computer-mediated instruction on L2 vocabulary acquisition. A medium effect size was found favoring computer-mediated instruction and other moderating variables were further investigated. What is noteworthy is that one of the variables she investigated was game-based learning. She found computer-mediated instructions without gaming elements ($d = 1.113$) were better than those with gaming elements ($d = 0.495$) in terms of L2 vocabulary learning. However, she did not report the result of Q test so it was

difficult to determine whether or not the difference between computer-mediated instructions without gaming elements and those with gaming elements was statistically significant. In addition, she did not investigate moderating variables regarding DGBL so her finding concluding that L2 vocabulary learning performed better without games than with the aid of games might be misleading. Therefore, the effects of DGBL have not been thoroughly explored in previous meta-analyses of second/foreign language learning. While Chiu's meta-analysis is quite recent, the current meta-analysis is complementary to her meta-analysis since the current one further investigated more potential moderating variables to provide a more comprehensive picture of the effects of DGBL in EFL contexts.

The Relationship Between Task-based Language Learning and DGBL

In the research field of task-based language learning, the use of tasks has recently drawn much attention from language educators (e.g., Adams, 2009; Ellis, 2009; Révész, 2009; Robinson & Gilabert, 2007). Teachers can use tasks to elicit students' language performance and know how to assist their learning (Nunan, 1989). Since language learning requires contexts, students can use languages to communicate in tasks. Furthermore, tasks serve as a medium to help researchers document how learners structure and restructure their interlanguage over time so that they can trace how learner language develops. The importance of tasks cannot be overemphasized in L2 teaching. Long (1991) claimed that previous teaching methods look similar and no teaching method is found the most effective. He broke the boundary among teaching methods, distinguishing "focus on form" from "focus on forms." The former draws students' attention to linguistic forms in a lesson with the primary focus on meaning and communication; the latter mainly focuses on only learning of discrete points of grammar. He, however, does not really describe how to "focus on form" but point out the importance of task designs in L2 teaching. A task is designed to provide learners with opportunities to primarily engage in meaningful contexts and to draw students' attention to linguistic form. Since the definition of a task in the past was inconsistent in the area of L2 teaching, Ellis (2003) organized previous researchers' definitions of tasks, establishing widely accepted criteria to define what a task should be. He established six criteria to define what a task is, including a task as a work plan, a focus on meaning, involving real-world language use, involving any of four language skills, engaging cognitive processes, and a clearly defined communicative outcome (Ellis, 2003).

Recent researchers have claimed that there are some relationships between task-based language learning and DGBL. Rapeepisarn, Wong, Fung, and Khine (2008) pointed out that different game genres are suitable for different learning objectives of language tasks. For example, learners in role play games can integrate different language skills to solve language problems they face in learning tasks. Franciosi (2011) further indicated that digital games with certain design features provide an optimal language learning situation for task participants to get involved in. A coding scheme of potential moderating variables in task characteristics was therefore developed based on the relationship between task-based language learning and DGBL (see Table 2).

Table 2: The Potential Moderating Variables in Different Task Characteristics

Task characteristics	Potential moderating variables
Task participant characteristics	<ul style="list-style-type: none"> • Learners' educational levels • Instructor's bias
Task design characteristics	<ul style="list-style-type: none"> • Internet facilities • Game types
Task involvement characteristics	<ul style="list-style-type: none"> • Treatment duration • Linguistic knowledge

The Use of Meta-analysis

In L2 research, it is not possible to determine whether certain interventions are beneficial to language learning simply based on a single study's statistical significance since results typically vary with different research contexts (Norris & Ortega, 2006). A meta-analysis is a systematic review to objectively decide whether certain intervention is effective or not and reach a firm conclusion by synthesizing data across studies (Wampold, Ahn, & Kim, 2000). In the issue of DGBL, to use digital games or not to use has been debated in L2 learning research since previous empirical studies yielded conflicting results. A meta-analysis is appropriately conducted to address this controversial issue of DGBL.

In the procedure of the meta-analysis, the first step is to calculate effect sizes from each study. Since more than one effect size in a single study inflate sample sizes, lose independence of data points, and cause distortion of standard error, only one effect size per study can be included for a meta-analysis (Borenstein, Hedges, Higgins, & Rothstein, 2009). The principle of "one study, one effect size" is therefore followed in this study. It is necessary to decide which comparison in each study to include in the meta-analysis. Since

this study attempts to explore the effects of DGBL, the comparison between the control group and the digital game treatment group in each study is extracted from their performance in posttests. The following widely used guideline of Cohen's d (Cohen, 1992) is adopted to interpret effect sizes in the meta-analysis: small effect = 0.2–0.5; medium effect = 0.5–0.8; large effect = 0.8 and up. As to how to produce average effect sizes from several comparisons between control and treatment groups, previous researchers have proposed some methods. Since variations exist among different studies, variations need to be taken into consideration when an effect size is extracted across studies. In meta-analyses, two statistical models are commonly used to address variation problems: the fixed-effect model and the random-effects model (Borenstein et al., 2009). Under the fixed-effect model, it is assumed that all studies are identical and there is only one true effect size for these studies. Any variation is attributable to sampling variability. In contrast, under the random-effects model, true effect sizes could vary from study to study and all studies are considered similar rather than identical. The variation results from heterogeneous factors. Because the two models produce slightly different results, to report the result from either model would be misleading (Li, 2010). Therefore, the results from the two models are reported to show a comprehensive picture of these studies.² To accurately interpret an average effect size, besides the power of effect sizes decided by Cohen's d , a 95% confidence interval (CI) is taken into consideration. The 95% CI expects that about 95% of the CIs constructed from different data sets include real average effect size and about 5% will fail to do so. A CI that does not include zero in the range indicates a 95% confidence in that the true effect size is included within this range. To examine which potential moderating variables are more influential than others, Q -tests are performed to determine whether the potential moderating variables investigated in this meta-analysis are significant moderators of effectiveness of feedback. To investigate whether a publication bias exists among these studies, Begg and Mazumdar's (1994) rank correlation test is carried out. If the publication bias exists among included studies in a meta-analysis, variances would significantly correlate with effect sizes (Borenstein et al., 2009). Finally the test for heterogeneity is conducted to detect the heterogeneity of the effect size obtained in this meta-analysis. To interpret the degree of the heterogeneity of effect size, the following guideline is adopted: $I^2 = 0\% \rightarrow$ no heterogeneity, $I^2 = 25\% \rightarrow$ low heterogeneity, $I^2 = 50\% \rightarrow$ moderate heterogeneity, and $I^2 = 75\% \rightarrow$ high heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003). In this article, professional meta-analysis software, Comprehensive Meta-Analysis (Borenstein et al., 2005), is performed to obtain statistical data for the meta-analysis.

Inclusion Criteria

Since this meta-analysis is conducted to investigate the effects of DGBL in EFL contexts, studies in which digital games are used or developed for educational purposes to learn English are selected for inclusion in this meta-analysis. Several inclusion criteria are established to decide which studies are appropriately included in the meta-analysis. First of all, selected studies must compare the effects of experimental groups (students who played educational digital games) with control groups (students who received traditional instructions) to show the treatment effects. Second, studies must be conducted in the EFL context where English is a foreign language, not as a second language, for daily communication. Third, since it is a quantitative meta-analysis, only quantitative studies which provide statistical data for effect size calculation are included. Finally, studies which assess students' learning effects from achievement tests in posttests are selected because the analysis is to explore the effects of DGBL on students' English learning performance. Studies which only investigate students' motivation or attitude toward DGBL are therefore excluded.

In a meta-analysis, the so-called apples and oranges problem forced the question of what made studies similar enough to be included and what made them different enough to be excluded (Norris & Ortega, 2006). With respect to second/foreign language studies, the most controversial problem affecting construct validity was whether studies were carried out in classroom or laboratory settings (Lyster & Saito, 2010). Only quasi-experimental classroom-based studies are therefore included, providing sufficient basis for comparison by means of meta-analysis.

The Literature Search and Coding

Studies to be included in the analysis were derived from six databases. These databases included: Chinese Periodical Index, Dissertation and Thesis Abstract System of Taiwan, IEEE Xplore, ERIC, Linguistics and Language Behavior Abstracts, and Google Scholar. The combination of the following keywords were used: (a) digital game based learning (DGBL), (b) computer assisted language learning (CALL), (c) technology enhanced language learning (TELL), (d) English as a foreign language (EFL), (e) digital games, (f) educational games, and (g) English education. In addition, the ancestry approach (Cooper, 1982) was used to locate studies by tracking reference sections of relevant studies which were already retrieved from the databases. Data collection was completed by the beginning

of April 2014. Finally, twenty-five studies were found to satisfy the inclusion criteria. Nine master's theses were retrieved from Dissertation and Thesis Abstract System of Taiwan. Five conference papers were derived from IEEE Xplore. Six journal papers were selected from ERIC. Three journal papers were found in Google Scholar. One journal paper and one conference paper were retrieved from references in the paper found through computer searches.

In the meta-analysis, there may be differences in the setting or practices or conditions of different studies that may influence the effects of DGBL and that conflating all studies without regard to these differences could result in a failure to detect their role. To help address this concern, several potential moderating variables in each kind of task characteristics are examined. The task participant characteristics include students' educational levels (elementary school levels vs. high school or college levels) and instructor's bias (possible presence of instructor's bias vs. absence of instructor's bias). The task design characteristics include Internet facilities (networked games vs. non-networked games) and game types (meaningful and engaging games vs. drill and practice games). The task involvement characteristics include treatment duration (short treatment duration vs. long treatment duration) and linguistic knowledge (declarative knowledge vs. procedural knowledge). These variables are described and discussed as follows.

In EFL contexts, language learners in high school (or college) levels normally have more pressures from taking tests than those in elementary schools (Chiu, 2013; Choi, 2008). Because of different levels of learning pressures from taking tests, the distinction was thus made between elementary school levels and high school (or college) levels in evaluating the effects of DGBL in terms of learners' educational levels. Among these twenty-five studies, fifteen studies recruited elementary school students as subjects whereas ten studies recruited high school or college students as subjects.

In CALI/CALL research, it has been debated whether or not the computer-based instruction (in experimental groups) and the traditional instruction (in control groups) should be designed by the same instructor (Chang & Lin, 2013; Clark & Leonard, 1985). If both groups are taught or guided by the same instructor, an instructor's bias might occur and have some confounding effects on research results. Among the twenty-five studies, sixteen studies possibly have the instructor's bias while other nine studies have not.

DGBL research has suggested that Internet facilities might promote more interactions among game players (T. Y. Liu & Chu, 2010; Suh, Kim, & Kim, 2010). Since interactions among second/foreign language learners are considered necessary for language acquisition

(Long, 1996), it is therefore particularly worthwhile to explore whether digital games equipped with Internet facilities are more effective for learning than those without Internet facilities (Liao et al., 2010). Therefore, digital games that were hosted on the Internet and provided Internet facilities were coded as networked games while digital games that were hosted only on the computer without the access to the Internet were coded as non-networked games. It was found that twelve of the twenty-five studies provided networked games and thirteen non-networked games.

In the research area of DGBL, whether action-based drill and practice educational games or meaningful and engaging games are effective has been debated (Kiili, 2005; Yip & Kwan, 2006). Some researchers claimed that drill and practice games, compared with meaningful and engaging games, may not lead to learning because players simply experiment with actions until they gain high scores (Kiili, 2005). Some researchers, on the contrary, argued that if action-based drill and practice games are designed well, this type of games will still be beneficial to players' learning (Yip & Kwan, 2006). As for the game types, drill and practice games were therefore distinguished from meaningful and engaging games in this meta-analysis. Operationalizations of drill and practice games as well as meaningful and engaging games were adopted from Kiili (2005). In drill and practice games, only story backgrounds are introduced at the beginning of the games and no storyline is further provided. After the introduction of the story backgrounds, players play the games by simply modifying actions until their scores improve. These behaviors are only based on trial and error. In contrast, meaningful and engaging games provide strong storylines and opportunities for players to explore phenomena, test hypotheses, and construct objects. Such educational games lead to direct experience with the game world. It was found that seventeen of the twenty-five studies used drill and practice games and eight used meaningful and engaging games.

Treatment duration is normally a significant variable in CALI/CALL research because of a novelty effect. Clark (1983) suggested when a treatment of a new medium lasted for a long period of time, the learning effect of technology use will decrease. For example, Chiu's (2013) meta-analysis showed that EFL learners who received CALI for less than one month learned vocabulary better than those who received instruction for more than one month. Therefore, to examine the effects of treatment duration, a distinction was drawn between short treatment duration and long treatment duration. Following the suggestions about the coding of treatment duration from Chiu (2013) and Li (2010), studies that reported the duration within one month were coded as short treatment duration while studies that

reported the duration for more than one month were coded as long treatment duration. In terms of the treatment duration, twelve studies that reported providing the treatment of DGBL within one month were coded as short treatment duration while thirteen studies that reported providing the treatment of DGBL for more than one month were coded as long treatment duration.

In foreign language learning research, learners' linguistic knowledge can be divided into declarative knowledge (i.e., knowing knowledge) and procedural knowledge (i.e., being able to apply the knowledge when needed) (J. R. Anderson, 1981). There has been a debate of whether declarative knowledge could be converted into procedural knowledge in the field of SLA (e.g., Krashen, 1985; McLaughlin, 1987). The distinction between declarative knowledge and procedural knowledge was thus investigated in this meta-analysis. The declarative knowledge was assessed to test whether learners knew certain linguistic knowledge. The procedural knowledge was assessed to test whether learners were able to apply certain linguistic knowledge in real language use. Among the twenty-five studies, fifteen studies assessed declarative knowledge while ten studies assessed procedural knowledge.

To investigate the effects of DGBL, students' language performance shown on immediate posttests was extracted from selected studies. Following Li (2010), a short-term immediate posttest was considered as an assessment given within one week after treatment. Thus, posttests which were given immediately after the treatment of DGBL or within one week after the treatment were considered as immediate posttests. Since not all studies provide the information of students' language performance on delayed posttests, it is not possible to examine the long-term learning effects shown on delayed posttests in this meta-analysis.

To obtain more reliable outcomes from coding, a trained research colleague was invited to code the studies. To check for accuracy, two coders coded each of the studies independently. Cohen kappa was performed to determine if there was agreement between two coders' judgments on the coding of moderating variables. To interpret the strength of agreement of Cohen kappa, the following guideline is followed: $\kappa < 0.00 \rightarrow$ poor, $\kappa = 0.00-0.20 \rightarrow$ slight, $\kappa = 0.21-0.40 \rightarrow$ fair, $\kappa = 0.41-0.60 \rightarrow$ moderate, $\kappa = 0.61-0.80 \rightarrow$ substantial, and $\kappa = 0.81-1.00 \rightarrow$ almost perfect (Landis & Koch, 1977). There was almost perfect agreement between the two coders' judgments on the coding of students' educational levels ($\kappa = 1.00$), the coding of instructor's bias ($\kappa = 0.89$), the coding of Internet facilities ($\kappa = 0.92$), the coding of game types ($\kappa = 0.87$), the coding of the treatment duration

($\kappa = 0.94$), and the coding of linguistic knowledge ($\kappa = 0.92$) in the initial coding. Finally, collaborative coding was conducted for the coding of the variables in which the two coders initially disagreed with each other.

To calculate an effect size (Cohen's d) for each study, three statistical data were extracted, including group mean scores (both control and experimental groups), standard deviations and sample sizes. The following equation from Cohen (1988) was mainly used. In the equation, M_1 and M_2 are control and experimental group mean scores respectively and σ is the standard deviation.

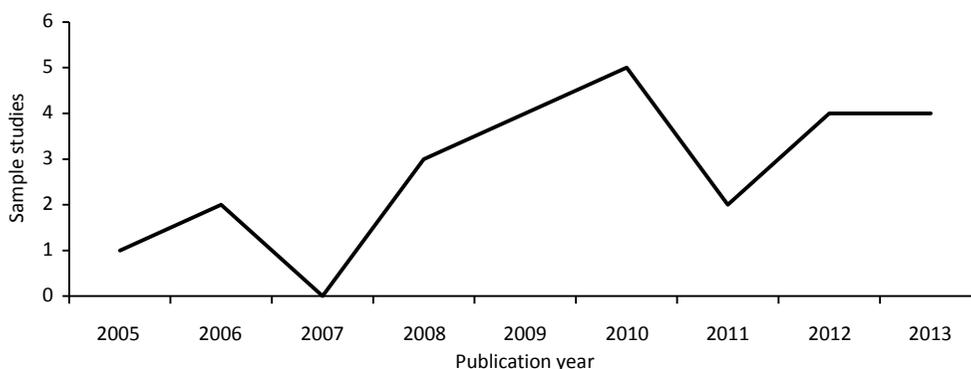
$$\frac{(n_1 - 1)\sigma_1 + (n_2 - 1)\sigma_2}{(n_1 - 1) + (n_2 - 1)} = \sigma_{pooled} \quad \frac{M_1 - M_2}{\sigma_{pooled}} = d$$

Results

Demographic Data

There were totally 1,908 students in the twenty-five selected studies. The frequency distribution of the included studies was displayed in Figure 1. As shown, there has been a gradual growth in the number of studies on DGBL since 2008.

Figure 1: Publication Frequency of Empirical Studies in DGBL



Overall Effectiveness

Table 3 demonstrated weighted effect sizes of DGBL. The result indicated that there was a medium positive effect size ($d = 0.695$) with the 95% CI of 0.598 to 0.791 under the fixed-effect model and also a medium positive effect size ($d = 0.777$) with the 95% CI

Table 3: Summary of DGBL Effect Size of Primary Studies

Studies	Cohen's <i>d</i>	Standard error	Lower limit	Upper limit
Aghlara & Tamjid (2011)	0.666	0.325	0.029	1.303
T. A. F. Anderson (2009)	0.267	0.266	-0.254	0.789
T. A. F. Anderson, Reynolds, Yeh, & Huang (2008)	0.024	0.372	-0.704	0.753
Barreira et al. (2012)	0.714	0.405	-0.079	1.507
Chen (2008)	0.485	0.226	0.042	0.929
Chiang (2013)	0.566	0.279	0.020	1.112
Chung (2012)	0.501	0.227	0.056	0.946
Fotouhi-Ghazvini, Earnshaw, Robison, & Excell (2009)	1.819	0.752	0.346	3.293
Huang (2009)	0.436	0.190	0.064	0.807
H. C. Hung, Young, & Lin (2013)	0.160	0.366	-0.557	0.877
Y. C. Hung (2009)	0.365	0.265	-0.154	0.884
Kim & Kim (2005)	0.563	0.186	0.198	0.927
T. Y. Liu & Chu (2010)	0.954	0.264	0.437	1.471
Lu (2010)	-0.069	0.258	-0.575	0.438
Niu (2011)	0.517	0.283	-0.037	1.072
Sadeghi & Dousti (2013)	1.440	0.351	0.752	2.127
Shyu (2006)	0.497	0.279	-0.050	1.044
Suh et al. (2010)	1.851	0.161	1.534	2.167
Sylvén & Sundqvist (2012)	1.111	0.310	0.503	1.720
Vahdat & Behbahani (2013)	2.063	0.391	1.296	2.830
Wang, Chen, & Shih (2012)	0.304	0.118	0.073	0.535
J. C. Yang, Chen, & Jeng (2010)	0.135	0.258	-0.371	0.642
J. C. Yang, Lin, Wu, & Chien (2008)	0.176	0.296	-0.404	0.756
Y. T. Yang (2010)	3.767	0.416	2.951	4.583
Yip & Kwan (2006)	1.265	0.220	0.834	1.696
Effect size (Cohen's <i>d</i>)				
Fixed-effect model	0.695	0.049	0.598	0.791
Random-effects model	0.777	0.140	0.504	1.051

of 0.504 to 1.051 under the random-effects model. Since the CIs did not include zero, the observed averaged effect sizes were statistically trustworthy. Moreover, Begg and Mazumdar's (1994) rank correlation test showed that the correlation between variances and effect sizes in these studies did not reach a statistically significant level ($z = 1.705, p > .05$). This suggested that the effect sizes obtained in this meta-analysis were not confounded by

a publication bias. In addition, the test for heterogeneity indicated that the effect size was highly heterogeneous ($I^2 = 86.381, p < .05$), which suggested that some potential moderating variables needed to be taken into account.

Task Participant Characteristics

Learners' educational levels

Table 4 showed effect sizes of DGBL for studies with different educational levels. In terms of studies which recruited elementary school students, the DGBL yielded large positive effect sizes under both the fixed-effect model ($d = 0.822$) and the random-effects model ($d = 0.818$). In terms of studies which recruited high school or college students, the DGBL yielded medium positive effect sizes under both the fixed-effect model ($d = 0.558$) and the random-effects model ($d = 0.701$). The difference in improvement between the control and experimental groups in these studies, taken together, reached statistical significance, as evidenced by the 95% CI. This indicated that the effect sizes for DGBL with different students' educational levels were trustworthy. There was a significant difference between DGBL with elementary school students and DGBL with high school or college students under the fixed-effect model ($Q = 7.184, p = 0.007$) but no significant difference was found under the random-effects model ($Q = 0.187, p = 0.666$).

Table 4: DGBL Effect Sizes for Learners' Educational Levels

	Elementary levels ($n = 15$)		High school and college levels ($n = 10$)	
	Fixed-effect model	Random-effects model	Fixed-effect model	Random-effects model
Effect size	0.822	0.818	0.558	0.701
Standard error	0.068	0.211	0.071	0.172
Variance	0.005	0.044	0.005	0.030
CI (95%)	Upper: 0.957	Upper: 1.231	Upper: 0.697	Upper: 1.038
	Lower: 0.688	Lower: 0.405	Lower: 0.419	Lower: 0.363

Instructor's bias

Table 5 demonstrated DGBL effect sizes for the possible presence and absence of instructor's bias separately. As to the studies with a possible instructor's bias, the DGBL yielded medium positive effect sizes under both the fixed-effect model ($d = 0.749$) and the random-effects model ($d = 0.753$). As to the studies without an instructor's bias, the DGBL yielded a medium positive effect size under the fixed-effect model ($d = 0.607$) but a large

Table 5: DGBL Effect Sizes for Instructor's Bias

	Possible presence of instructor's bias ($n = 16$)		Absence of instructor's bias ($n = 9$)	
	Fixed-effect model	Random-effects model	Fixed-effect model	Random-effects model
Effect size	0.749	0.753	0.607	0.823
Standard error	0.063	0.188	0.080	0.218
Variance	0.004	0.035	0.006	0.048
CI (95%)	Upper: 0.872	Upper: 1.121	Upper: 0.763	Upper: 1.252
	Lower: 0.627	Lower: 0.384	Lower: 0.450	Lower: 0.395

effect size under the random-effects model ($d = 0.823$). Moreover, a significant difference between studies with a possible instructor's bias and studies without an instructor's bias was not found under the fixed-effect model ($Q = 1.982, p = 0.159$). No significant difference was also found between them under the random-effects model ($Q = 0.060, p = 0.806$).

Task Design Characteristics

Internet facilities

Table 6 presented effect sizes of DGBL for studies with different Internet facilities. In terms of studies with networked games, the DGBL yielded a large positive effect size under the fixed-effect model ($d = 0.844$) but a medium positive effect size under the random-effects model ($d = 0.719$). In terms of studies with non-networked games, the DGBL yielded a medium positive effect size under the fixed-effect model ($d = 0.533$) but a large effect size under the random-effects model ($d = 0.861$). In addition, since the 95% CI did not include zero, the observed effect sizes were considered statistically trustworthy. Furthermore, the Q -test revealed that there was a significant difference between networked game and non-networked game studies under the fixed-effect model ($Q = 9.937, p = 0.002$) but no significant difference was found under the random-effects model ($Q = 0.255, p = 0.614$).

Table 6: DGBL Effect Sizes for Internet Facilities

	Networked game ($n = 12$)		Non-networked game ($n = 13$)	
	Fixed-effect model	Random-effects model	Fixed-effect model	Random-effects model
Effect size	0.844	0.719	0.533	0.861
Standard error	0.068	0.173	0.071	0.224
Variance	0.005	0.030	0.005	0.050
CI (95%)	Upper: 0.978	Upper: 1.057	Upper: 0.673	Upper: 1.300
	Lower: 0.710	Lower: 0.380	Lower: 0.394	Lower: 0.423

Game types

Table 7 demonstrated the differential effects of DGBL types. Drill and practice games yielded medium positive effect sizes under both the fixed-effect model ($d = 0.511$) and the random-effects model ($d = 0.620$). On the contrary, meaningful and engaging games yielded large positive effect sizes under both the fixed-effect model ($d = 1.101$) and the random-effects model ($d = 1.117$). Since the 95% CI did not include zero, the observed effect sizes were considered statistically trustworthy. The Q -test further revealed that there was a significant difference between the two game types under the fixed-effect model ($Q = 30.705, p = 0.000$) but no significant difference was found under the random-effects model ($Q = 2.804, p = 0.094$).

Table 7: DGBL Effect Sizes for Game Types

	Drill and practice game ($n = 17$)		Meaningful and engaging game ($n = 8$)	
	Fixed-effect model	Random-effects model	Fixed-effect model	Random-effects model
Effect size	0.511	0.620	1.101	1.117
Standard error	0.059	0.153	0.088	0.255
Variance	0.004	0.023	0.008	0.065
CI (95%)	Upper: 0.628 Lower: 0.395	Upper: 0.919 Lower: 0.321	Upper: 1.274 Lower: 0.928	Upper: 1.616 Lower: 0.618

Task Involvement Characteristics

Treatment duration

Table 8 contained effect sizes of DGBL for studies with different treatment duration. In terms of studies of short treatment duration, the DGBL yielded small positive effect sizes under both the fixed-effect model ($d = 0.425$) and the random-effects model ($d = 0.425$). In terms of studies of long treatment duration, the DGBL yielded large positive effect sizes under both the fixed-effect model ($d = 0.888$) and the random-effects model ($d = 1.092$). In addition, since the 95% CI did not include zero, the observed effect sizes were considered statistically trustworthy. Furthermore, the Q -test revealed that there were significant differences between studies of short treatment duration and those of long treatment duration under the fixed-effect model ($Q = 21.411, p = 0.000$) and the random-effects model ($Q = 7.110, p = 0.008$).

Table 8: DGBL Effect Sizes for Treatment Duration

	Short treatment duration (<i>n</i> = 12)		Long treatment duration (<i>n</i> = 13)	
	Fixed-effect model	Random-effects model	Fixed-effect model	Random-effects model
Effect size	0.425	0.425	0.888	1.092
Standard error	0.076	0.076	0.065	0.238
Variance	0.006	0.006	0.004	0.057
CI (95%)	Upper: 0.575 Lower: 0.276	Upper: 0.575 Lower: 0.276	Upper: 1.014 Lower: 0.761	Upper: 1.558 Lower: 0.625

Linguistic knowledge

Table 9 demonstrated DGBL effect sizes for different linguistic knowledge. The DGBL for declarative knowledge yielded small positive effect sizes under both the fixed-effect model ($d = 0.432$) and the random-effects model ($d = 0.489$). The DGBL for procedural knowledge yielded large positive effect sizes under both the fixed-effect model ($d = 1.153$) and the random-effects model ($d = 1.126$). The difference in improvement between the control and experimental groups for these DGBL studies, taken together, reached statistical significance, as evidenced by the 95% CI. This indicated that the DGBL effect sizes for either declarative or procedural knowledge obtained were reliable. Significant differences between the learning effects for declarative and procedural knowledge were found under both the fixed-effect model ($Q = 49.503, p = 0.000$) and the random-effects model ($Q = 4.714, p = 0.030$).

Table 9: DGBL Effect Sizes for Linguistics Knowledge

	Declarative knowledge (<i>n</i> = 15)		Procedural knowledge (<i>n</i> = 10)	
	Fixed-effect model	Random-effects model	Fixed-effect model	Random-effects model
Effect size	0.432	0.489	1.153	1.126
Standard error	0.062	0.099	0.082	0.276
Variance	0.004	0.010	0.007	0.076
CI (95%)	Upper: 0.554 Lower: 0.311	Upper: 0.682 Lower: 0.295	Upper: 1.313 Lower: 0.993	Upper: 1.668 Lower: 0.584

Discussion

This meta-analysis generally showed that a medium positive effect size was found to favor DGBL. The meta-analysis therefore suggested DGBL to be effective for the acquisition of English as a foreign language. In addition, publication bias appeared to be

negligible in this meta-analysis. Finally, the overall effect size obtained in this meta-analysis was heterogeneous. It was therefore suggested that other moderating variables need to be taken into account when investigating the effects of DGBL. Six potential moderating variables in three task characteristics (i.e., task participant characteristics: students' educational levels and instructor's bias; task design characteristics: Internet facilities and game types; task involvement characteristics: treatment duration and linguistic knowledge) were therefore investigated and certain variables were found to be more influential than others in determining the DGBL effectiveness. The variables investigated in this meta-analysis are discussed in the following sections.

Task Participant Characteristics

The task participant characteristics contain two potential moderating variables: learners' educational levels and instructor's bias. In terms of learners' educational levels, DGBL was effective for both elementary school students and high school (or college) students although no significant difference was found between them. The possible reason might be that English learning has long been exam-oriented in the EFL context. Since washback can be positive or negative (Bailey, 1996), playing digital games might be able to minimize the difference between teaching and testing, promoting positive washback effects for EFL learners for taking exams (Weigle & Jensen, 1997). The positive washback effects might increase the effects of DGBL regardless of students' educational levels. In terms of the instructor's bias, no significant difference was found between studies with a possible instructor's bias and those without an instructor's bias. The instructor's bias therefore does not serve as a moderator variable in determining the DGBL effects. Since Clark and Leonard (1985) claimed that an instructor's bias might influence the effects of computer-based instruction, the finding reported in this meta-analysis does not support their claim. The reason could be due to the learners' strong interests in the gaming elements in the treatment. Learners might be too attracted by digital games to notice instructors' intervention in the treatment.

Task Design Characteristics

In task design characteristics, two potential moderating variables, Internet facilities and game types, are included. As to Internet facilities, although most networked games (about 8 out of 12 studies) offer multiplayer interactions to promote learners' interactions with their

counterparts which are necessary to foreign language learning (Long, 1996), the networked games are not found to be more effective for learners' acquisition of English than the non-networked games. This might be because it is already sufficient for learning to occur in EFL contexts when language learners interact with game characters or figures to negotiate meaning in digital games. An investigation into game types revealed that meaningful and engaging games yielded a large effect size while drill and practice games yielded a medium effect size. Nonetheless, no significant difference was found between the two game types. The possibility might be that EFL learners already get used to the learning style in drill and practice games because of their previous learning experience in traditional EFL classrooms. Another possibility might be that meaningful and engaging games may not show effects in the short term. Yip and Kwan (2006) found that students who showed interests in drill and practice games in class might not be willing to play the games on their own after class. Therefore, if the effects of meaningful and engaging games show up later, beyond the window of immediate posttests, this meta-analysis would have missed these longer-term effects shown in delayed posttests. As mentioned earlier, delayed posttest data were excluded since it was not possible to get the data for all of the studies. Thus another goal for future research could be to include delayed posttest data for meta-analysis to see if meaningful and engaging games produce any more trustworthy statistical results.

Task Involvement Characteristics

Overall, DGBL was found to be effective for EFL learners. Two potential moderating variables in task involvement characteristics were further investigated and found to be influential in determining the effectiveness of DGBL, which is the major finding in the current meta-analysis. The two moderating variables are treatment duration and linguistic knowledge. It is found that conflating the effects of treatment duration or linguistic knowledge ($d = 0.777$) overestimates the effects of short treatment duration ($d = 0.425$) or declarative knowledge ($d = 0.489$) and underestimates the effects of long treatment duration ($d = 1.092$) or procedural knowledge ($d = 1.126$). In what follows, the two significant moderator variables were discussed. As to treatment duration, students who received the long treatment duration of DGBL learn English better than those who received the short treatment duration. This may indicate that EFL learners can be motivated by the application of digital games in the beginning and their motivation can be maintained for a long time (Schank & Neaman, 2001). Since learners' long immersion in language environments is necessary for the success of language learning (Genesee, 1994; Johnson & Swain, 1997),

digital games provide a simulated context of language use and serve as an immersion environment to increase the chance of further exposure to target language input and opportunities for output in a long period of time. For example, in T. Y. Liu and Chu's (2010) study, a campus environment was simulated for learners to use English in a library, a laboratory, an auditorium and other school buildings. In another example from Sadeghi and Dousti (2013), a family environment was simulated for learners to introduce their family members to friends in English. In terms of linguistic knowledge, DGBL was found to be more effective for learners' acquisition of procedural knowledge than those of declarative knowledge. Chiu's (2013) meta-analysis showed L2 learners performed better without digital games than with the help of digital games in terms of vocabulary learning. Vocabulary knowledge, however, can be declarative (i.e., knowing vocabulary) or procedural (i.e., being able to apply the vocabulary when needed) (J. R. Anderson, 1981). For example, in Yip and Kwan's (2006) study, two vocabulary game Websites were used to increase learners' declarative and procedural knowledge of English vocabulary. Not only an explanation of a word's meaning but also the use of the word in a sentence or a passage are provided in the vocabulary games. Their result showed that learning with the vocabulary games was more effective than traditional learning without vocabulary games. The current meta-analysis, therefore, furthers their finding, indicating that digital games could facilitate more acquisition of procedural knowledge than acquisition of declarative knowledge. This finding suggests that digital games provide a platform for learners to apply their knowledge. Swain (1985) found that even after many years of exposure to the target language, learners' language output still contained errors because learners did not have an opportunity to practice linguistic forms in classrooms. Digital games thus provide opportunities for learners to use their knowledge to produce language.

In SLA, whether declarative knowledge can be converted into procedural knowledge has been keenly debated over the years. Krashen (1985) distinguishes "acquisition" from "learning," claiming that they are two separate processes. He maintains that knowledge of the second language acquired in natural or meaningful communication and learned in classroom instruction or activities cannot become unified. In other words, the learned declarative knowledge cannot be converted into acquired procedural knowledge. His claim, nevertheless, has been criticized for a perceived lack of scientific rigor. J. R. Anderson (1993) and McLaughlin (1987), on the contrary, support that declarative knowledge can be converted into procedural knowledge from controlled toward automatic processing. L2 learners first resort to controlled processing by means of explicit instruction of declarative

knowledge; through practice over time, declarative knowledge could become automatized procedural knowledge. This meta-analysis therefore justifies J. R. Anderson's and McLaughlin's claims, suggesting that digital games may serve as a useful learning task to automatize procedural knowledge. The digital games could help EFL learners convert declarative knowledge into procedural knowledge particularly when they are immersed in DGBL environments for more than one month.

Theoretical and Pedagogical Implications

The significant finding of treatment duration and linguistic knowledge in the task involvement characteristics may have theoretical and pedagogical implications. As for theoretical implications, the significant effects of long treatment duration and procedural knowledge in DGBL serve as direct evidence supporting Craik and Lockhart's (1972) levels of processing theory. They claimed that there are two levels of processing (shallow processing and deep processing) and the deeper the level of processing, the easier the information could be recalled. Since the involvements in the long treatment duration and procedural knowledge require learners' deep processing in DGBL tasks, language learners are allowed to take time and could easily recall the information they receive from digital games to achieve success in language learning. As regards pedagogical implications, future digital game designers could develop games which encourage EFL learners to apply their linguistic knowledge for communicative purposes and attract learners to get immersed in language learning environments provided by digital games for a long period of time.

Limitations and Suggestions for Future Studies

There are some limitations in the current meta-analysis as follows. First of all, because of limited access to the databases, all master's theses included in this meta-analysis were conducted in Taiwan. It is suggested that future researchers get access to databases to derive unpublished studies (e.g., master's theses) in other countries of the Asia-Pacific region. Secondly, since the current meta-analysis investigated the general effectiveness of digital game-based English learning with analyses of limited number of moderator variables, more variables should be taken into account to advance the understanding of DGBL in EFL contexts. Furthermore, this study investigated only EFL learners' cognitive outcomes. It is suggested that future studies should be conducted to synthesize studies to explore EFL learners' affective outcomes. Thirdly, since only the short-term learning effects of DGBL in

immediate posttests are examined in this study, it is worthwhile to investigate whether the effects of the moderating variables investigated in this meta-analysis could have long-term learning effects shown in delayed posttests.

Conclusion

In this meta-analysis, an overall substantial medium positive effect size for DGBL was found. Furthermore, large positive effect sizes were found for digital games designed to maintain learners' long-term engagement and to help automatize their procedural knowledge. This suggests that the evolution of research on DGBL in EFL contexts has reached a stage that it is no longer useful or interesting to ask whether DGBL makes a difference. The results in this meta-analysis suggest that rather than conflating all sorts of DGBL practices to seek an answer to the broad question of whether digital games work, it is more useful to acknowledge the evidence now that some sorts work and some sorts do not and to concentrate future research efforts on learning more about the difference to advance the understanding of DGBL in EFL contexts.

Notes

1. Different from EFL learning which normally occurs in classrooms, SLA is a systematic study of how people learn a second language or a language other than their mother tongue inside or outside classrooms (Ellis, 1997).
2. Although both statistical models are reported in this meta-analysis, it should be noteworthy that the effect sizes under the random-effects model are much more trustworthy than those under the fixed-effect model. This is because the various DGBL studies that are included in this meta-analysis have not been designed with the intention of being identical to other such studies (in contrast, say, to studies conducted with the intention of duplicating previous studies). Thus, it might be inappropriate in this case to assume the true effect size is exactly the same in all the studies, and so accordingly, the random-effects model should be considered as the more suitable approach for determining the effects of DGBL.

Acknowledgments

The author would like to thank the anonymous reviewers for their valuable comments on the draft versions of this article. This research has been supported and funded by the grant from China Foundation for the Promotion of Education and Culture.

References

- Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, 21(3), 199–226. doi: 10.1080/09588220802090246
- Adams, R. (2009). Recent publications on task-based language teaching: A review. *International Journal of Applied Linguistics*, 19(3), 339–355. doi: 10.1111/j.1473-4192.2009.00239.x
- Aghlara, L., & Tamjid, N. H. (2011). The effect of digital games on Iranian children's vocabulary retention in foreign language acquisition. *Procedia – Social and Behavioral Sciences*, 29, 552–560. doi: 10.1016/j.sbspro.2011.11.275
- Anderson, J. R. (1981). *Cognitive skills and their acquisition*. Hillsdale, NJ: Lawrence Erlbaum.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum.
- Anderson, T. A. F. (2009). *Design of an English vocabulary acquisition system with mini-games for elementary school students in Taiwan* (Unpublished master's thesis). National Central University, Taiwan.
- Anderson, T. A. F., Reynolds, B. L., Yeh, X. P., & Huang, G. Z. (2008). Video games in the English as a foreign language classroom. In M. Eisenberg, Kinshuk, M. Chang, & R. McGreal (Eds.), *The second IEEE international conference on digital game and intelligent toy enhanced learning* (pp. 188–192). doi: 10.1109/DIGITEL.2008.39
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257–279. doi: 10.1177/026553229601300303
- Barreira, J., Bessa, M., Pereira, L. C., Adão, T., Peres, E., & Magalhães, L. (2012). *MOW: Augmented reality game to learn words in different languages*. Retrieved from <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6263236>
- Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50(4), 1088–1101.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2005). *Comprehensive meta-analysis* [Computer software]. Englewood, NJ: Biostat.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: John Wiley & Sons.
- Chang, M. M., & Lin, M. C. (2013). Strategy-oriented Web-based English instruction — A meta-analysis. *Australasian Journal of Educational Technology*, 29(2), 203–216. Retrieved from <http://ascilite.org.au/ajet/submission/index.php/AJET/article/view/67/52>
- Chen, H. Q. (2008). 國小學童英語音素能力的電腦化訓練環境 [A computer-based training environment for children's phonemic awareness] (Unpublished master's thesis). National Taiwan Normal University, Taiwan.
- Chiang, Y. J. (2013). *A study on the effectiveness of a smart phone gaming application on foreign language learning: A case of English verb-noun collocation learning* (Unpublished master's thesis). National Taiwan Normal University, Taiwan.

- Chiu, Y. H. (2013). Computer-assisted second language vocabulary instruction: A meta-analysis. *British Journal of Educational Technology*, 44(2), E52–E56. doi: 10.1111/j.1467-8535.2012.01342.x
- Choi, I. C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25(1), 39–62. doi:10.1177/0265532207083744
- Christensen, E., Merrill, P., & Yanchar, S. (2007). Second language vocabulary acquisition using a diglot reader or a computer-based drill and practice program. *Computer Assisted Language Learning*, 20(1), 67–77. doi: 10.1080/09588220601118511
- Chung, L. Y. (2012). Virtual reality in college English curriculum: Case study of integrating second life in freshman English course. In L. Barolli, T. Enokido, F. Xhafa, & M. Takizawa (Eds.), *Proceedings: 26th IEEE international conference on advanced information networking and applications workshops* (pp. 250–253). doi: 10.1109/WAINA.2012.198
- Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53(4), 445–459. doi: 10.3102/00346543053004445
- Clark, R. E., & Leonard, S. (1985, March). *Computer research confounding*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, U.S.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. doi: 10.1037/0033-2909.112.1.155
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers and Education*, 59(2), 661–686. doi: 10.1016/j.compedu.2012.03.004
- Cooper, H. M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of Educational Research*, 52(2), 291–302. doi: 10.3102/00346543052002291
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. doi: 10.1016/S0022-5371(72)80001-X
- Ellis, R. (1997). *Second language acquisition*. Oxford, England: Oxford University Press.
- Ellis, R. (2003). Tasks in SLA and language pedagogy. In R. Ellis, *Task-based language learning and teaching* (pp. 1–35). Oxford, England: Oxford University Press.
- Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, 30(4), 474–509. doi: 10.1093/applin/amp042
- Felix, U. (2005). Analysing recent CALL effectiveness research — Towards a common agenda. *Computer Assisted Language Learning*, 18(1–2), 1–32. doi: 10.1080/09588220500132274
- Ferdig, R. E. (2006). Assessing technologies for teaching and learning: Understanding the importance of technological pedagogical content knowledge. *British Journal of Educational Technology*, 37(5), 749–760. doi: 10.1111/j.1467-8535.2006.00559.x

- Fotouhi-Ghazvini, F., Earnshaw, R., Robison, D., & Excell, P. (2009). The MOBO city: A mobile game package for technical language learning. *International Journal of Interactive Mobile Technologies*, 3(2), 19–24. doi: 10.3991/ijim.v3i2.757
- Franciosi, S. J. (2011). A comparison of computer game and language-learning task design using flow theory. *CALL-EJ*, 12(1), 11–25. Retrieved from http://callej.org/journal/12-1/Franciosi_2011.pdf
- Garrett, N. (1987). A psycholinguistic perspective on grammar and CALL. In W. F. Smith (Ed.), *Modern media in foreign language education: Theory and implementation* (pp. 169–196). Lincolnwood, IL: National Textbook.
- Genesee, F. (1994). Integrating language and content instruction. In R. M. Bostwick (Ed.), *Second language acquisition through content based study: An introduction to immersion* (pp. 44–54). Numazu, Japan: Katoh.
- Grgurović, M., Chapelle, C. A., & Shelley, M. C. (2013). A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL*, 25(2), 165–198. doi: 10.1017/S0958344013000013
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557–560. doi: 10.1136/bmj.327.7414.557
- Huang, C. J. (2009). 融合科學與英語學習之數位遊戲模式發展及成效研究 [A functional digital game can help children learn science and English] (Unpublished master's thesis). National Taipei University of Education, Taiwan.
- Hung, H. C., Young, S. S. C., & Lin, C. P. (2013). No student left behind: A collaborative and competitive game-based learning environment to reduce the achievement gap of EFL students in Taiwan. *Technology, Pedagogy and Education*. doi: 10.1080/1475939X.2013.822412
- Hung, Y. C. (2009). 遊戲式英語句型學習對國小學生學習成效之影響 [The effect of game-assisted English sentence patterns learning on elementary students' learning achievement] (Unpublished master's thesis). National Taiwan Normal University, Taiwan.
- Johnson, R. K., & Swain, M. (1997). *Immersion education: International perspectives*. Cambridge, England: Cambridge University Press.
- Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education*, 8(1), 13–24. doi: 10.1016/j.iheduc.2004.12.001
- Kim, H. S., & Kim, S. B. (2005). An integrated course based on educational games. In H. Selvaraj & P. K. Srimani (Eds.), *Proceedings: ITCC 2005. International conference on information technology: Coding and computing* (Vol. 1, pp. 436–441). doi: 10.1109/ITCC.2005.81
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. London, England: Longman.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.

- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*(2), 309–365. doi: 10.1111/j.1467-9922.2010.00561.x
- Liao, Y. K., Chang, L. H., & Chang, C. C. (2010). Game-based learning vs. traditional instruction: A meta-analysis of thirty-eight studies from Taiwan. In C. Maddux (Eds.), *Research highlights in information technology and teacher education 2010* (pp. 1491–1498). Chesapeake, VA: Society for Information Technology and Teacher Education.
- Lin, W. C., Huang, H. T., & Liou, H. C. (2013). The effects of text-based SCMC on SLA: A meta analysis. *Language Learning and Technology, 17*(2), 123–142. Retrieved from <http://ilt.msu.edu/issues/june2013/linetal.pdf>
- Liu, M., Moore, Z., Graham, L., & Lee, S. (2003). A look at the research on computer-based technology use in second language learning: A review of the literature from 1990–2000. *Journal of Research on Technology in Education, 34*(3), 250–273. doi: 10.1080/15391523.2002.10782348
- Liu, T. Y., & Chu, Y. L. (2010). Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers and Education, 55*(2), 630–643. doi: 10.1016/j.compedu.2010.02.023
- Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. B. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). Amsterdam, the Netherlands: John Benjamins.
- Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). San Diego, CA: Academic Press.
- Lu, C. Y. (2010). 悅趣化合作學習融入國中英語補救教學研究 [The study of game-based cooperative learning into English remedial instruction for junior high school] (Unpublished master's thesis). National Chiayi University, Taiwan.
- Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition, 32*(2), 265–302. doi: 10.1017/S0272263109990520
- McLaughlin, B. (1987). *Theories of second language learning*. London, England: Edward Arnold.
- Niu, Y. R. (2011). 線上社交遊戲策略對國小學生英語單字學習之影響 [Effects of online social game strategy on elementary students' English vocabulary learning] (Unpublished master's thesis). National Taiwan Normal University, Taiwan.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*(3), 417–528. doi: 10.1111/0023-8333.00136
- Norris, J. M., & Ortega, L. (2006). The value and practice of research synthesis for language learning and teaching. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 3–50). Amsterdam, the Netherlands: John Benjamins.

- Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge, England: Cambridge University Press.
- Oberg, A. (2011). Comparison of the effectiveness of a CALL-based approach and a card-based approach to vocabulary acquisition and retention. *CALICO Journal*, 29(1), 118–144.
- Perez, M. M., Noortgate, W. V. D., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41(3), 720–739. doi: 10.1016/j.system.2013.07.013
- Rapeepisarn, K., Wong, K. W., Fung, C. C., & Khine, M. S. (2008). The relationship between game genres, learning techniques and learning styles in educational computer games. In Z. Pan, X. Zhang, A. El Rhalibi, W. Woo, & Y. Li (Eds.), *Technologies for e-learning and digital entertainment* (pp. 497–508). Berlin, Germany: Springer-Verlag.
- Reinders, H. (2012). *Digital games in language learning and teaching*. Basingstoke, England: Palgrave Macmillan.
- Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, 31(3), 437–470. doi: 10.1017/S0272263109090366
- Robinson, P., & Gilabert, R. (2007). Task complexity, the cognition hypothesis and second language learning and performance. *International Review of Applied Linguistics in Language Teaching*, 45(3), 161–176. doi: 10.1515/iral.2007.007
- Sadeghi, K., & Dousti, M. (2013). The effect of length of exposure to CALL technology on young Iranian EFL learners' grammar gain. *English Language Teaching*, 6(2), 14–26. doi: 10.5539/elt.v6n2p14
- Schank, R., & Neaman, A. (2001). Motivation and failure in educational simulation design. In K. D. Forbus & P. J. Feltovich (Eds.), *Smart machines in education: The coming revolution in educational technology* (pp. 37–69). Cambridge, MA: AAAI Press/The MIT Press.
- Shyu, J. Y. (2006). 運用多媒體遊戲提升兒童英語認字拼字與學習態度之研究——以桃園縣某國小為例 [Effects of implementing the computer games on English vocabulary learning and attitudes] (Unpublished master's thesis). National Taipei University of Education, Taiwan.
- Suh, S., Kim, S. W., & Kim, N. J. (2010). Effectiveness of MMORPG-based instruction in elementary English education in Korea. *Journal of Computer Assisted Learning*, 26(5), 370–378. doi: 10.1111/j.1365-2729.2010.00353.x
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley, MA: Newbury House.
- Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, 24(3), 302–321. doi: 10.1017/S095834401200016X

- Vahdat, S., & Behbahani, A. R. (2013). The effect of video games on Iranian EFL learners' vocabulary learning. *The Reading Matrix*, 13(1), 61–71. Retrieved from http://www.readingmatrix.com/articles/april_2013/vahdat_behbahani.pdf
- Wampold, B. E., Ahn, H. N., & Kim, D. M. (2000). Meta-analysis in the social sciences: A useful way to make sense of a series of findings from a large number of studies. *Asia Pacific Education Review*, 1(1), 67–74. doi: 10.1007/BF03026147
- Wang, H. C., Chen, C. T., & Shih, H. (2012). *The use of game-based pilot system in English learning for low-achieving students*. Retrieved from <ftp://163.14.2.14/scu/network/tanet2012/node/proceedings/paper/N-210-1.pdf>
- Warschauer, M., & Kern, R. (2000). *Network-based language teaching: Concepts and practice*. Cambridge, England: Cambridge University Press.
- Weigle, S. C., & Jensen, L. (1997). Issues in assessment for content-based instruction. In M. A. Snow & D. M. Brinton (Eds.), *The content-based classroom: Perspectives on integrating language and content* (pp. 201–212). White Plains, NY: Longman.
- Wouters, P., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers and Education*, 60(1), 412–425. doi: 10.1016/j.compedu.2012.07.018
- Yang, J. C., Chen, C. H., & Jeng, M. C. (2010). Integrating video-capture virtual reality technology into a physically interactive learning environment for English learning. *Computers and Education*, 55(3), 1346–1356. doi: 10.1016/j.compedu.2010.06.005
- Yang, J. C., Lin, Y. L., Wu, J. J., & Chien, K. H. (2008). Design and evaluation of a physical interactive learning environment for English learning. In M. Eisenberg, Kinshuk, M. Chang, & R. McGreal (Eds.), *The second IEEE international conference on digital game and intelligent toy enhanced learning* (pp. 36–40). doi: 10.1109/DIGITEL.2008.40
- Yang, Y. T. (2010). 數位遊戲式英語教學之學習成效與滿意度研究——以國小二年級學生為例 [The effect and satisfaction of digital games based English teaching — A case study of the second year elementary school students] (Unpublished master's thesis). Kainan University, Taiwan.
- Yip, F. W. M., & Kwan, A. C. M. (2006). Online vocabulary games as a tool for teaching and learning English vocabulary. *Educational Media International*, 43(3), 233–249. doi: 10.1080/09523980600641445
- Zhao, Y. (2003). Recent developments in technology and language learning: A literature review and meta-analysis. *CALICO Journal*, 21(1), 7–27.

英語為外語環境下數位遊戲式學習任務效益之後設分析

高千文

摘要

本研究採用後設分析，探討數位遊戲式學習在英語為外語環境下之效益為何。針對 25 篇研究進行分析後，發現中度且正面的效果量（固定效果模式： $d = 0.695$, $p < .05$ ；隨機效果模式： $d = 0.777$, $p < .05$ ）。其結果顯示，在英語為外語環境裏，數位遊戲式學習相較於傳統教學（如：文法翻譯法或聽說教學法）有效。本研究根據任務導向語言學習與數位遊戲式學習的關係發展一套編碼系統，進一步分析數位遊戲式學習的效益為何。該編碼系統包含三種任務特徵，而這三種特徵共有六項潛在調節變項。在這些潛在調節變項中，任務投入特徵裏的實驗時間及語言知識對數位遊戲式學習最有顯著效果。根據結果，本後設分析研究建議，以英語為外語學習者接受較長時間的數位遊戲式學習，對其程序性語言知識的發展助益較大。未來研究應探討其他潛在調節變項對數位遊戲式學習有何效益，得以更深入了解該領域的全貌。

關鍵詞：數位遊戲式學習；後設分析；英語為外語

