Brain, Digital, & Learning

pISSN : 2384-2474
eISSN : 2586-7490

RESEARCH ARTICLE

A Meta-Analysis of Data-Driven Learning (DDL) in EFL/ESL Settings

Kyung-Ho Yoon[1], Dong Ju Lee[2]*
[1]Gonghang Middle School
[2]Korea National University of Education

*Corresponding Author: Dong Ju Lee, maydjlee@knue.ac.kr

ABSTRACT

This research is on the effect size (ES) of corpus-based data-driven learning (DDL) in EFL/ESL, which can show the effectiveness/efficiency of this particular instructional method relative to those of others. This meta-analysis methodology in instructed second language acquisition (ISLA) has been developed and established by Norris and Ortega (2000), Plonsky and Oswald (2014), etc. The initial search results in Education Resources Information Center (ERIC) reveal 5,165 research articles, but the final number is reduced to 46 (54 unique samples) after applying step-by-step inclusion criteria. The weighted mean ES (Hedges's *g*) between the comparison and experimental groups is 1.11 (*SE*: 0.13), which is large. The weighted mean ES between the pre-tests and immediate post-tests is 1.81 (*SE*: 0.16), which is large, too. The delayed post-test analyses are also conducted. In addition, the present meta-analysis investigates the ESs influenced by the seven moderator variables (MVs). The above-mentioned results as a whole indicate that the ES of corpus-based DDL in EFL/ESL is much larger than that of the overall ISLA, and DDL may have some specific MV subgroups where it is more effective/efficient. These results suggest that more detailed research be conducted on DDL which looks promising as a whole.

Key words: Meta-analysis, effect size, data-driven learning (DDL), corpus, concordance

Brain, Digital, & Learning
2024, Vol. 14, No. 2, 283-304
https://doi.org/10.31216/BDL.20240017

Received: June 04, 2024
Revised: June 17, 2024
Accepted: June 19, 2024

Introduction

The instructional applications of corpus linguistics could be said to be rooted from Sinclair (1987), Biber (1988), Johns (1990, 1991), etc. Following these initiators, a large number of scholars (e.g., Boulton, 2010; Cobb, 1999; Francis, 1993; Hoey, 2005; Hunston & Francis, 2000) have been developing corpus linguistics and its pedagogical applications both theoretically and practically. Such scholars pointed out that lexis and grammar cannot be separated but should be consolidated in English-as-a-foreign-language/English-as-a-second-language (EFL/ESL) instructions, where a corpus can be a strong instructional instrument providing authentic lexico-grammatical patterns.

While corpus-based instructional methods including data-driven learning (DDL) have a history of more than 30 years in the Western world, its applications in East Asia have been experimented and developed for the last 20 years or so. For instance, Tono (2002), Sripicharn (2003), and Lee

(2007) applied the established DDL methods to Japan, Thailand, and Korea, respectively, in accordance with the diverse local EFL situations. Especially, Lee (2007) proved that corpus-based DDL can be effective for low- or intermediate-level EFL learners as in Korea if the soft-version of DDL is utilized instead of the hard-version. These initiators in East Asia seem to have researched and tried to disseminate DDL in their countries in order to overcome the chronic problems in the unfavorable EFL situations where English instructions are implemented mostly without native speaking teachers or even authentic materials.

The present study intends to analyze/synthesize this history of corpus-based DDL in EFL/ESL. As will be described more in detail in the chapter Literature Review, there have been actually several previous research syntheses or meta-analyses on DDL. However, most of them were qualitative research syntheses or narrow-focused meta-analyses, which are not appropriate for measuring the overall learning outcomes from DDL. As a result, there are only two previous meta-analyses (e.g., Boulton & Cobb, 2017; Lee et al., 2019) left as directly related to the current meta-analysis. However, the former meta-analysis where the latest primary study included was published in 2014 is rather outdated in this fast-changing field of DDL, and it just reported the simple mean effect size instead of the weighted mean effect size for estimating the overall effect of DDL. The latter meta-analysis is more sophisticated in terms of statistical instruments, but it researched only on vocabulary learning in DDL. In addition, Lee et al.'s (2019) meta-analysis did not conduct the within-group analysis but performed only the between-group analysis whereas there are more within-group primary studies than between-group ones on corpus-based DDL.

Consequently, a new comprehensive meta-analysis is needed to synthesize these DDL instructions developed in diverse regions and with various pedagogical methods resulting from the entire history of corpus-based DDL. While utilizing on the above-mentioned two meta-analyses' empirical findings and academic achievement, the present study will try to improve the research methods in order to show the overall learning outcomes from DDL more precisely and comprehensively.

In relation to this broad research topic on DDL, the particular research purpose of this study is to investigate the effect sizes of corpus-based DDL in EFL/ESL, which can show the effectiveness and efficiency of this particular instructional method relative to those of others. In short, this research is a meta-analysis whose methodology has been developed and established by Norris and Ortega (2000), Plonsky and Oswald (2014), and so on. This meta-analysis methodology naturally leads to the following two research questions:

1. What is the overall effect size of corpus-based data-driven learning (DDL) in English as a foreign language (EFL) or English as a second language (ESL)?

2. What is the influence of empirically/theoretically-identified moderator variables on the effect sizes of corpus-based data-driven learning (DDL) in English as a foreign language (EFL) or English as a second language (ESL)?

Literature Review

Since this meta-analysis intends to study the academic field of corpus-based DDL, starting with a brief introduction to corpora themselves would be appropriate. Boulton and Vyatkina (2021) noted that corpora emerged in their modern form in the 1960s as large collections of electronic texts were designed to represent an area of language use.

Compared to traditional linguistics, corpus linguistics is an empirical approach totally relying on the corpus evidences. Hunston (2022) noted that “corpus linguistics is an approach to the study of language that involves collecting large quantities of naturally occurring language and using specialised software that manipulates that language to obtain information about frequencies, co-occurrences and meanings” (p. 1). A concordancer searches corpora and presents formatted outputs such as concordance lines (usually providing with several words before/after the searched word) that enable patterning to be observed. In brief, corpus linguistics is for the empirical generalization of language features based on the observation of language use in the real world.

Corpus linguistics has brought about important changes in linguistics and applied linguistics since the 1970s (Hunston, 2022). Corpus research enabled variations of a language differentiated by place, time, and context to be studied in more detail than before. Also, corpus linguistics has given lexico-grammar and phraseology a more important role in the linguistic scene.

Pedagogically, a corpus can help produce authentic textbooks, learner dictionaries, and other instructional materials containing more details about how words and expressions are used in the real world. Also, a learner corpus can help identify the language features that are difficult or easy for a specific group of learners and the stages/processes in acquiring a language (Lee, 2007).

The early use of corpora for language learners was to develop authentic reference materials such as dictionaries and grammar books (Hunston, 2022; O’Keeffe et al., 2007). Then, learner corpora were compiled to improve the understanding of language acquisition, which could lead to the development of higher-quality instructional materials (Lee, 2007). At last, the pedagogic effectiveness of the use of corpora by language learners themselves was identified, which is the subject of this meta-analysis. Johns (1990, 1991) used corpus data to encourage college students to discover which phraseology was preferred in specific discourses, and this kind of “data-driven learning” (DDL) has become part of the mainstream of applied corpus research, which can lead to meeting learners’ needs and promoting learner autonomy in relation to discovery learning.

Moving on to the meta-analysis methodology itself, it was developed to synthesize the effects of multiple experiments in such academic disciplines as medicine, business, education, etc. (see Hwang, 2020). Following this original tradition, the meta-analysis in instructed second language acquisition (ISLA) can be said to have originated from Norris and Ortega (2000). Their meta-analysis method deducted the weighted mean effect size from primary quantitative studies based on the descriptive statistics including the mean, standard deviation, and sample size. After the Norris and Ortega’s seminal work, diverse scholars such as Plonsky and Oswald (2014) or Plonsky and Brown (2015) have been refining the meta-analysis methodology in ISLA.

In the field of DDL, there are eight research syntheses or meta-analyses found by the authors after excluding conference papers. However, three out of these eight journal articles (e.g., Boulton & Vyatkina, 2021; Chen & Flowerdew, 2018; Paquot & Plonsky, 2017) are just research syntheses without calculating the weighted mean effect sizes although they provide a deep insight into this field of DDL. Apart from the research syntheses, Durrant’s (2014) study is a meta-analysis, but it focused only on the correlation between learner knowledge and frequency data of collocations and did not synthesize the effects of overall DDL. In addition, Mizumoto and Chujo's (2015) meta-analysis researched only on Japanese learners of English. Therefore, there are only three meta-analysis studies (e.g., Boulton & Cobb, 2017; Cobb & Boulton, 2015; Lee et al., 2019) left as the direct previous studies to guide this meta-analysis. Among these three meta-analyses, however, Cobb and Boulton's (2015) one was just a preliminary meta-analysis for the same scholars' next comprehensive work (Boulton & Cobb, 2017). Consequently, the present study refers more often to the remaining two meta-analyses on L2-related DDL (e.g., Boulton & Cobb, 2017; Lee et al., 2019).

Boulton and Cobb's (2017) and Lee et al.'s (2019) meta-analyses provide the present study with some invaluable benchmarks. Boulton and Cobb (2017) selected 64 studies that included 88 unique samples, and Lee et al. (2019) researched 29 vocabulary-related studies that had 38 unique samples in them. Both of the meta-analyses used Hedges's *g* (Hedges & Olkin, 1985) instead of Cohen's *d* (Cohen, 1988) to measure effect sizes as the current study does. In Boulton and Cobb's (2017) research, the mean effect size of between-group (post-tests of the comparison and experimental groups) analysis was 0.95 while that of Lee et al.'s (2019) was 0.74. The mean effect size of delayed post-test (between-group) analysis was 0.11 in Boulton and Cobb's (2017) study whereas that of Lee et al.'s (2019) was 0.64.

Contrary to Mizumoto and Chujo (2015) who conducted only the within-group analysis, Lee et al.'s (2019) meta-analysis excluded the within-group analysis and instead performed only the between-group analysis. As opposed to these two studies, Boulton and Cobb (2017) conducted both the between- and within-group analyses. Their mean effect size of within-group (the pre- and immediate post-tests of an experimental group) analysis was 1.50 while that for the pre- and delayed post-tests was 1.36.

In accordance with Boulton and Cobb's (2017) meta-analysis, the present study conducts both the between- and within-group analyses. Also, this meta-analysis researches the entire scope of corpus-based DDL, not limited to a specific field such as vocabulary in DDL, which was the only focus of Lee et al.'s (2019). Furthermore, the current study differentiates itself more from Boulton and Cobb's (2017), because this meta-analysis focuses on EFL/ESL excluding other L2s (Spanish and mixed) than English. In addition, the present study uses the weighted mean effect size although Boulton and Cobb (2017) used the simple mean of individual effect sizes, the latter of which is quite problematic in terms of the statistical accuracy of estimation.

Method

Samples for the Meta-Analysis

The current meta-analysis is based on the search results from Education Resources Information Center (ERIC), which has been used by most of the meta-analyses on ISLA. Plonsky and Brown (2015) indicated that, among 80 L2-related meta-analyses they investigated, 70 studies reported their selection of specific databases: ERIC was used by 81% of them, which is much greater than the ratio (49%) of the second mostly utilized database, Linguistics and Language Behavior Abstracts (LLBA). The other frequently-used databases were PsycINFO (41%), ProQuest Dissertations and Theses (39%), Web of Science (14%), and so on.

Although the present meta-analysis does not intend to select only the most frequently-used database (ERIC), the library in the authors' institution does not subscribe LLBA, PsycINFO, or Web of Science, which the authors cannot access individually. Consequently, these databases are simply excluded from this meta-analysis. Concerning ProQuest Dissertations and Theses, the current study does not search such a database because this meta-analysis intends to focus only on published peer-reviewed journal articles excluding unpublished doctoral dissertations.

ERIC was accessed via ProQuest in July, 2023, using such search keywords as *corpus/corpora, data driven/data-driven, DDL,* and *concordanc**—the wildcard character * made the search include all the related keywords such as *concordance, concordancer, concordancers, concordancing,* etc.—with the screening criteria of 'peer reviewed'.

Regarding the search period, the present meta-analysis searched throughout the entire history of corpus-based DDL. However, the initial search period is not the same as the resultant period of primary studies included in this meta-analysis because there has been an increasing number of primary quantitative studies that are more rigorous and robust to satisfy the inclusion criteria for a meta-analysis. As a result, the earliest primary study included in the current meta-analysis was published as late as in 1996, not in 1991 when the first quantitative research article on DDL (i.e., Stevens, 1991) was published.

Instrument and Procedure

The results of initial search were 5,165 primary studies in ERIC (via ProQuest) with the above- mentioned keywords included in anywhere throughout their texts. Among these 5,165 research articles, 73 quantitative studies on corpus-based DDL in EFL/ESL were left after excluding other categories as follow:

(a) The term 'corpus' referring to other objects than the corpus in linguistics such as the corpus callosum in a human brain; The term 'data-driven' being used differently from the context of DDL such as 'data-driven decision making'; The term 'concordance' being used differently from the context of DDL as the everyday meaning or a statistical term;

(b) Corpora being used for other academic disciplines than applied linguistics such as sociology; Corpora being utilized for other pedagogies than EFL/ESL ones such as the Chinese language or first language (L1) English pedagogies;

(c) Theoretical/conceptual explanations being provided instead of empirical research on DDL; Also, position papers, secondary studies, research syntheses, and meta-analyses different from primary research;

(d) A simple technical comparison of features/characteristics between different corpora; A technical introduction/explanation on corpus-related technology or software;

(e) Corpus-based material/test/curriculum development;

(f) A corpus-assisted analysis on textbooks, tests, interlanguage, or the language of a specific social group;

(g) Practical implementation guides or training materials for teachers or students instead of experiments;

(h) No experimental measurement reported on cognitive learning outcomes as in qualitative research or questionnaires.

The next stage was to investigate a primary experimental study's eligibility for a meta-analysis. First, some quantitative studies were excluded due to their experimental design issues: (a) the sampling problems such as participants' self-selection of groups, (b) the use of a control group instead of a comparison group, (c) the shortage of DDL due to a great amount of non-DDL instructions included for the experimental group, and so on. In addition, this meta-analysis excluded the primary studies where corpora were used as a reference tool during the test (writing, translation, error correction, etc.). This exclusion was for measuring the learning outcomes themselves of DDL instruction without any kind of external assistance. Lastly, this meta-analysis selected just 46 studies that reported the effect sizes themselves, or the inferential statistics such as *F*-statistic, or most preferably, the descriptive statistics including the mean, standard deviation, and sample size, with which the effect sizes can be calculated (see References for the list of these 46 quantitative studies finally included).

The number of unique samples is 54, not 46, however, because some studies have two or three unique samples in the same study. Among these 54 unique samples, 34 ones conducted between-group experiments comparing an experimental group to a comparison group, and 46 ones reported within-group experiments comparing a post-test result to a pre-test one.

With the 46 primary quantitative studies finally included, the authors tried to identify potential moderator variables that seemed to influence on the effects of corpus-based DDL in EFL/ESL. Investigating the moderators identified in the two previous meta-analyses (e.g., Boulton & Cobb, 2017; Lee et al., 2019) and considering potential moderators' theoretical and empirical foundations, the authors developed a coding book systematically through multiple stages in order to determine the moderator variables for this meta-analysis. The finally adopted moderator variables are (a) language area focus, (b) L2 proficiency, (c) instructional status/age, (d) region, (e) version of DDL, (f) type of corpus, and (g) intervention duration.

Data Analysis

According to Borenstein et al. (2009), Hedges's *g* (unbiased Cohen's *d*) (Hedges & Olkin, 1985) is a better measurement than Cohen's *d* (Cohen, 1988) to estimate the effect size of a small-size sample, the latter of which usually shows an upward bias. This superiority of Hedges's *g* for small-size samples was also indicated by Hedges and Olkin (1985) themselves with a specific benchmark (samples with $n < 50$). Therefore, the current meta-analysis adopts Hedges's *g* instead of Cohen's *d* since 14 out of 34 unique samples (between-group) have an experimental group whose sample size is 25 or less.

Regarding a specific formula for the spreadsheet (Microsoft Excel) calculation of Hedges's *g*, the present study utilizes the one in the book written by Hwang (2020), the required inputs of which are the mean, standard deviation, and sample size. In addition, a meta-analysis computer program is used to double-check the calculation results of effect sizes: Comprehensive Meta-Analysis (CMA), V4 (Professional Version) (Borenstein et al., 2022).

In other words, the current study uses the same formula as was utilized in Lee et al.'s (2019) meta-analysis in order to estimate the pooled standard deviations used for calculating individual effect sizes (Hedges's *g*) in between-group analyses (see Borenstein et al., 2009, p. 22, for the formula used in the present study). On the contrary, Boulton and Cobb (2017) used a simpler formula that did not consider the difference between the sample sizes of two compared groups, which is problematic in terms of statistical accuracy.

However, the current meta-analysis calculates the effect sizes for within-group contrasts with the same formula as is used for between-group contrasts, which is in line with Boulton and Cobb's (2017) approach. This approach is not statistically precise, though, since there must be a correlation between the pre- and post-tests for the same group (within-group). Instead, the pooled standard deviations to be used for calculating the individual effect sizes in within- group contrasts should include the correlation between the standard deviations of pre- and post-tests as follows (see Hwang, 2020, and Lee et al., 2019):

$$SD_{Gain} = \sqrt{SD_{Pre}^2 + SD_{Post}^2 - 2 \times Correlation_{Pre\&Post} \times SD_{Pre} \times SD_{Post}}$$

Unfortunately, there are actually no primary studies that report the correlation between the pre- and post-test results except for a few studies that provide with all the test scores of individual participants. As a result, just using the same formula as that for between-group contrasts is the best possible option as of now. When looking into the formula above, the two formulae for within- and between-group contrasts become similar if the correlation between pre- and post-tests is around .50, which seems not to be too different from the possible true correlation.

Even after manually calculating the individual effect sizes based on the descriptive statistics, there are still some effect sizes missing because some primary quantitative studies only reported inferential statistics such as *F*-statistic. In such cases, the *F*-statistic can be converted to Cohen's *d* by applying the following formula if the sample sizes are unequal between the comparison and experimental groups (for diverse formula to convert other statistical values to Cohen's *d*, see Borenstein et al., 2009, and Lipsey & Wilson, 2001):

$$\sqrt{\frac{F(n_1 + n_2)}{n_1 n_2}}$$

However, there are so diverse statistical values in the primary studies without fully reporting the mean, standard deviation, and sample size, so this meta-analysis converts such statistical values as the Mann-Whitney *U* test, chi-square, partial eta squared, and Pearson correlation coefficient (*r*) to Cohen's *d* by utilizing on the academic webpage *Computation of effect sizes* (Lenhard & Lenhard, 2016).

Lastly, in order to convert Cohen's *d* to Hedges's *g*, the following bias correction factor is used (see Borenstein et al., 2009, and Hedges & Olkin, 1985):

$$\left(1 - \frac{3}{4(n_1 + n_2) - 9}\right) d$$

In this way, the present study calculates 34 individual effect sizes of between-group unique samples and 46 individual effect sizes of within-group ones from the 54 unique samples in total.

With all the individual effect sizes as calculated above, the present meta-analysis estimates the weighted mean effect sizes (Hedges's *g*) of all the unique samples either for the between- or within-group contrasts by the meta-analysis program, CMA. Also, the weighted mean effect sizes for delayed post-tests are calculated with CMA both for the 12 between- and 13 within-group unique samples that reported delayed post-test results. CMA is also utilized for conducting the forest plot and the funnel plot analyses. The forest plot helps investigate how much the confidence intervals of all the effect sizes overlap with each other. The funnel plot examines the potential publication bias that could have distorted the results of a meta-analysis, which is not included in this short article, though.

The final stage is to estimate the influences of moderator variables on the effect sizes of subgroups classified by the coding book of this meta-analysis. For this subgroup analysis, the statistical significance is examined by the heterogeneity analysis employing the Q-value between the subgroups, which is also conducted with CMA.

Regarding the interpretation of effect sizes, the present meta-analysis adopts Plonsky and Oswald's (2014) benchmarks for L2 research instead of Cohen's (1988) traditional/general ones which are 0.20 (small), 0.50 (medium), and 0.80 (large). Plonsky and Oswald (2014) claimed that the field-specific benchmarks for L2-related research should be 0.40 (small), 0.70 (medium), and 1.00 (large) for between-group contrasts in terms of Cohen's d, which shows a highly similar value to Hedges's g (unbiased Cohen's d) used for this study. These benchmarks are in accordance with the fact that the mean effect size of overall L2-related research is 0.69 for Cohen's d, which is highly close to their benchmark for medium-size (0.70) (see Plonsky & Oswald, 2014, for various mean effect sizes of diverse social sciences). For effect sizes in within-group analysis, Plonsky and Oswald (2014) argued that the field-specific benchmarks for L2-related research should be 0.60 (small), 1.00 (medium), and 1.40 (large) in terms of Cohen's d.

Concerning the benchmark for a statistical significance of subgroup comparison, the present meta-analysis adopts the value '$p < .10$' instead of '$p < .05$', following the overall meta-analysis tradition (see Borenstein et al., 2009, for this unique benchmark of $p < .10$ specifically applied for the subgroup analysis instead of $p < .05$ used in most of the other cases).

Results

The Overall Effect Size of Between-Group Unique Samples

The weighted mean effect size calculated by the random-effects model (see Borenstein et al., 2010) with all the 34 between-group unique samples is 1.11, considered 'large' according to Plonsky and Oswald's (2014) benchmarks for between-group contrasts: 0.40 (small), 0.70 (medium), and 1.00 (large). The weighted mean effect size is presented at the bottom in Table 1 below, along with the individual effect sizes of each between-group unique sample, sorted low to high.

Table 1. The weighted mean effect size and the effect sizes of 34 between-group (immediate post-tests) unique samples

ID	Study Name (Unique Sample ID)	Hedges's *g*	*SE*	*p*-value
40	Lee & Lee, 2015	-0.88	0.27	0.00
45-2	Smart, 2014 (45-2)	-0.27	0.34	0.43
24	Yahya et al., 2020	-0.21	0.29	0.48
2	Shi, 2015	0.09	0.28	0.76
12	Boulton, 2010	0.37	0.18	0.04
6	Poole, 2012	0.43	0.45	0.35
39	Karras, 2015	0.45	0.20	0.03
38	Satake, 2022	0.61	0.27	0.02
43	Zahran, 2019	0.65	0.25	0.01
11-1	Lay & Yavuz, 2020a (11-1)	0.68	0.38	0.07
32-2	Bardovi-Harlig et al., 2017 (32-2)	0.80	0.39	0.04
46	Moon & Oh, 2017	0.90	0.31	0.00
19	Kilimci, 2017	0.92	0.29	0.00
37	Lee et al., 2017	0.93	0.13	0.00
18	Çakır & Özer, 2020	1.01	0.27	0.00
42	Kartal & Yangineksi, 2018	1.03	0.27	0.00
17	Tsai, 2021	1.04	0.21	0.00
26-3	Rezaee et al., 2015 (26-3)	1.07	0.27	0.00
32-1	Bardovi-Harlig et al., 2017 (32-1)	1.13	0.38	0.00
22	Gan et al., 1996	1.26	0.31	0.00
13	Alsuhaibani, 2022	1.32	0.23	0.00
31	Uçar & Yükselir, 2015	1.33	0.39	0.00
26-1	Rezaee et al., 2015 (26-1)	1.37	0.28	0.00
45-1	Smart, 2014 (45-1)	1.39	0.38	0.00
10	Daskalovska, 2015	1.54	0.33	0.00
35	Gordani, 2013	1.56	0.27	0.00
8	Suriyapee & Pongpairoj, 2022	1.61	0.29	0.00
26-2	Rezaee et al., 2015 (26-2)	1.72	0.30	0.00
5	Mirzaei et al., 2015	1.80	0.33	0.00
1	Kazaz, 2020	1.92	0.27	0.00
29	Mansoory & Jafarpour, 2014	2.36	0.40	0.00
33	Barabadi & Khajavi, 2017	2.38	0.34	0.00
3	Razaghi et al., 2022	2.90	0.45	0.00
44	Ashouri et al., 2014	3.32	0.40	0.00
	Weighted Mean Effect Size	1.11	0.13	0.00
	Mean of Prediction Interval	1.11		

Based on the statistics in Table 1, Fig. 1 graphically presents the same effect sizes along with the 95% confidence intervals. The weighted (pooled) mean effect size is displayed by the black (closed) diamond at the bottom, which shows that there is a 95% probability that the true weighted mean effect size falls between the lower limit of 0.86 and the upper limit of 1.36. Differently from this range of the precision of the estimate on past events, the capped red (gray) line below the black diamond is the prediction interval, which shows that there is a 95% chance that the effect size of a new experiment in the same field may fall anywhere between -0.31 and 2.53.

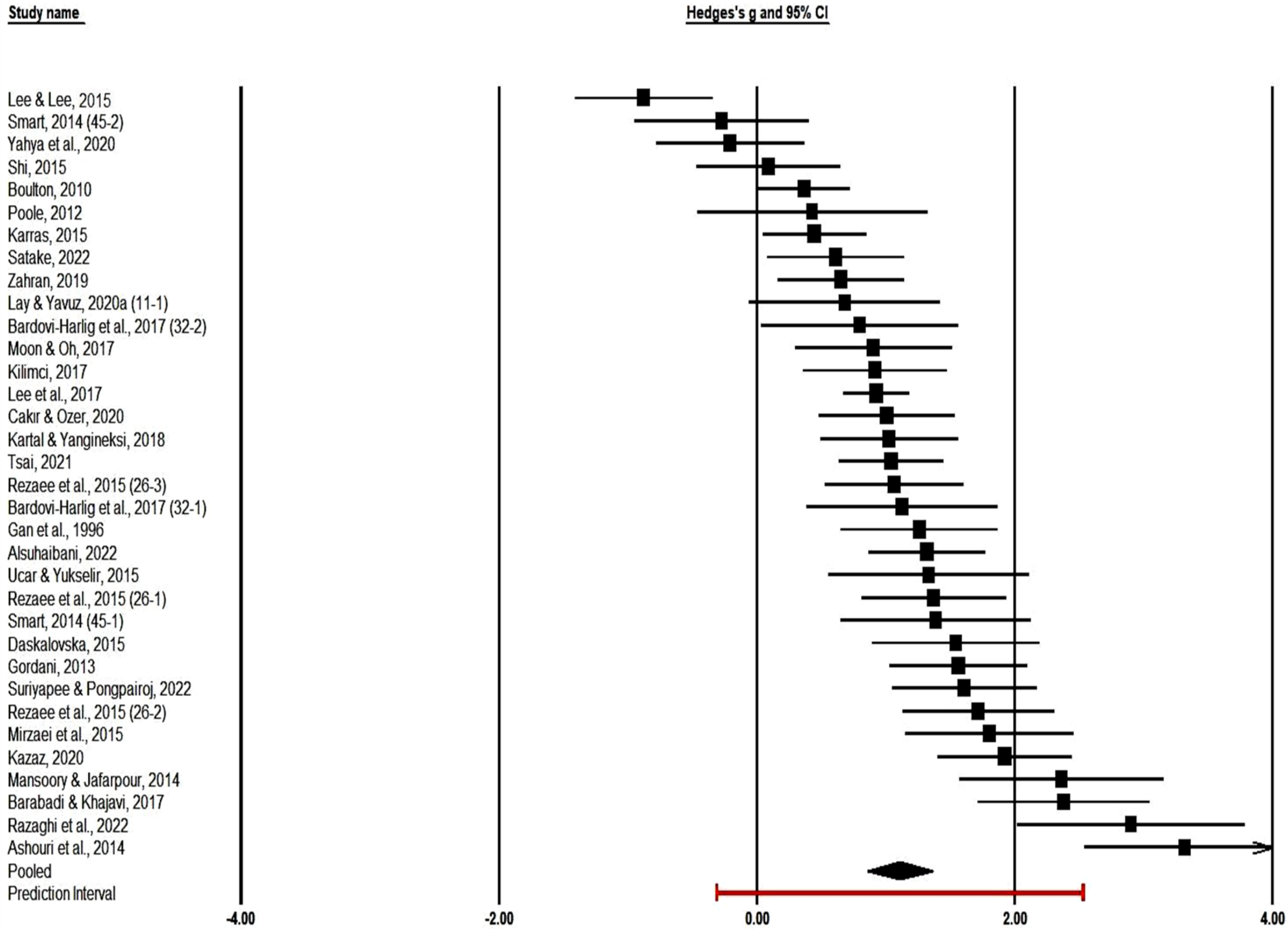

Fig. 1. The forest plot of effect sizes with 95% confidence intervals for 34 between-group (immediate post-tests) unique samples

As illustrated in Fig. 1, the individual effect sizes of 34 between-group (immediate post-tests) unique samples are all positive except for the first three unique samples. There is a huge difference between them, however, with the individual effect sizes ranging from -0.88 to 3.32, which is somewhat problematic. Actually, the lower limits of the six unique samples' effect sizes are even below zero. To the contrary, the last individual effect size is more than 3.00, which might be winsorized in some meta-analyses such as Boulton and Cobb's (2017) research (see Lipsey & Wilson, 2001, for the winsorization).

The weighted mean effect size calculated with all the 12 delayed-post-test between-group unique samples is 0.83, which is considered 'medium'. This weighted mean effect size is shown at the bottom in Table 2 below, along with the effect sizes of each unique sample, which are sorted low to high.

Table 2. The weighted mean effect size and the effect sizes of 12 delayed-post-test between-group unique samples

ID	Study Name (Unique Sample ID)	Hedges's *g*	*SE*	*p*-value
40	Lee & Lee, 2015	-0.86	0.27	0.00
45-2	Smart, 2014 (45-2)	0.17	0.34	0.63
38	Satake, 2022	0.27	0.27	0.32
12	Boulton, 2010	0.37	0.18	0.04
46	Moon & Oh, 2017	0.70	0.30	0.02
19	Kilimci, 2017	0.93	0.29	0.00
26-1	Rezaee et al., 2015 (26-1)	0.94	0.27	0.00
26-3	Rezaee et al., 2015 (26-3)	1.05	0.27	0.00
26-2	Rezaee et al., 2015 (26-2)	1.39	0.28	0.00
10	Daskalovska, 2015	1.52	0.33	0.00
45-1	Smart, 2014 (45-1)	1.76	0.40	0.00
1	Kazaz, 2020	1.87	0.26	0.00
	Weighted Mean Effect Size	0.83	0.22	0.00
	Mean of Prediction Interval	0.83		

The Overall Effect Size of Within-Group Unique Samples

Relative to the between-group analysis above, 46 within-group unique samples show even a larger difference between their effect sizes ranging from -0.70 to 16.17 (cf. the second largest one: 4.95) as illustrated in Table 3 below. Also in line with the large effect size of between-group ones, the weighted mean effect size of within-group unique samples is 1.81 (estimated by the random- effects model), which is considered 'large' in terms of Plonsky and Oswald's (2014) benchmarks for within-group contrasts: 0.60 (small), 1.00 (medium), and 1.40 (large). The effect size (1.81) of the within-group analysis is in line with and naturally larger than that (1.11) of the between-group analysis since it is hard for any kind of instruction not to produce any learning outcomes.

Table 3. The weighted mean effect size and the effect sizes of 46 within-group (pre- and immediate post-tests) unique samples (continued)

ID	Study Name (Unique Sample ID)	Hedges's *g*	*SE*	*p*-value
2	Shi, 2015	-0.70	0.29	0.02
4	Gaskell & Cobb, 2004	-0.06	0.31	0.85
11-2	Lay & Yavuz, 2020a (11-2)	-0.05	0.39	0.90
45-2	Smart, 2014 (45-2)	0.45	0.36	0.21
25-2	Lin & Lee, 2019 (25-2)	0.67	0.33	0.04
28	Lin, 2023	0.67	0.23	0.00
12	Boulton, 2010	0.73	0.18	0.00
6	Poole, 2012	0.87	0.47	0.07
32-2	Bardovi-Harlig et al., 2017 (32-2)	0.95	0.35	0.01
27	Lay & Yavuz, 2020b	0.95	0.27	0.00
11-1	Lay & Yavuz, 2020a (11-1)	1.01	0.42	0.02
25-1	Lin & Lee, 2019 (25-1)	1.03	0.36	0.00
9	Chen, 2017	1.11	0.31	0.00
15	Lin, 2021	1.15	0.16	0.00
14	Yoon & Jo, 2014	1.15	0.68	0.09
36	Girgin, 2019	1.18	0.20	0.00

Table 3. The weighted mean effect size and the effect sizes of 46 within-group (pre- and immediate post-tests) unique samples

ID	Study Name (Unique Sample ID)	Hedges's *g*	*SE*	*p*-value
18	Çakır & Özer, 2020	1.25	0.27	0.00
39	Karras, 2015	1.30	0.20	0.00
13	Alsuhaibani, 2022	1.47	0.23	0.00
5	Mirzaei et al., 2015	1.52	0.32	0.00
32-1	Bardovi-Harlig et al., 2017 (32-1)	1.53	0.31	0.00
24	Yahya et al., 2020	1.56	0.34	0.00
22	Gan et al., 1996	1.59	0.33	0.00
16-1	Cotos, 2014 (16-1)	1.62	0.40	0.00
17	Tsai, 2021	1.66	0.23	0.00
45-1	Smart, 2014 (45-1)	1.71	0.41	0.00
40	Lee & Lee, 2015	1.83	0.32	0.00
10	Daskalovska, 2015	1.89	0.37	0.00
31	Uçar & Yükselir, 2015	1.94	0.43	0.00
16-2	Cotos, 2014 (16-2)	1.99	0.44	0.00
34	Boontam, 2022	2.33	0.33	0.00
37	Lee et al., 2017	2.35	0.16	0.00
20	Cobb, 1997	2.37	0.54	0.00
21-1	Ruivivar & Lapierre, 2018 (21-1)	2.41	0.51	0.00
7	Liou et al., 2006	2.44	0.33	0.00
29	Mansoory & Jafarpour, 2014	2.58	0.41	0.00
41	Inpanich & Somphong, 2022	2.62	0.35	0.00
21-2	Ruivivar & Lapierre, 2018 (21-2)	2.67	0.60	0.00
43	Zahran, 2019	2.94	0.35	0.00
30	Hou, 2014	3.02	0.32	0.00
44	Ashouri et al., 2014	3.32	0.40	0.00
35	Gordani, 2013	3.54	0.38	0.00
23	Yeh et al., 2007	3.58	0.52	0.00
8	Suriyapee & Pongpairoj, 2022	4.13	0.44	0.00
1	Kazaz, 2020	4.95	0.44	0.00
3	Razaghi et al., 2022	16.17	1.83	0.00
	Weighted Mean Effect Size	1.81	0.16	0.00
	Mean of Prediction Interval	1.81		

Based on the statistics in Table 3 above, Fig. 2 graphically illustrates the same results along with the 95% confidence intervals of effect sizes.

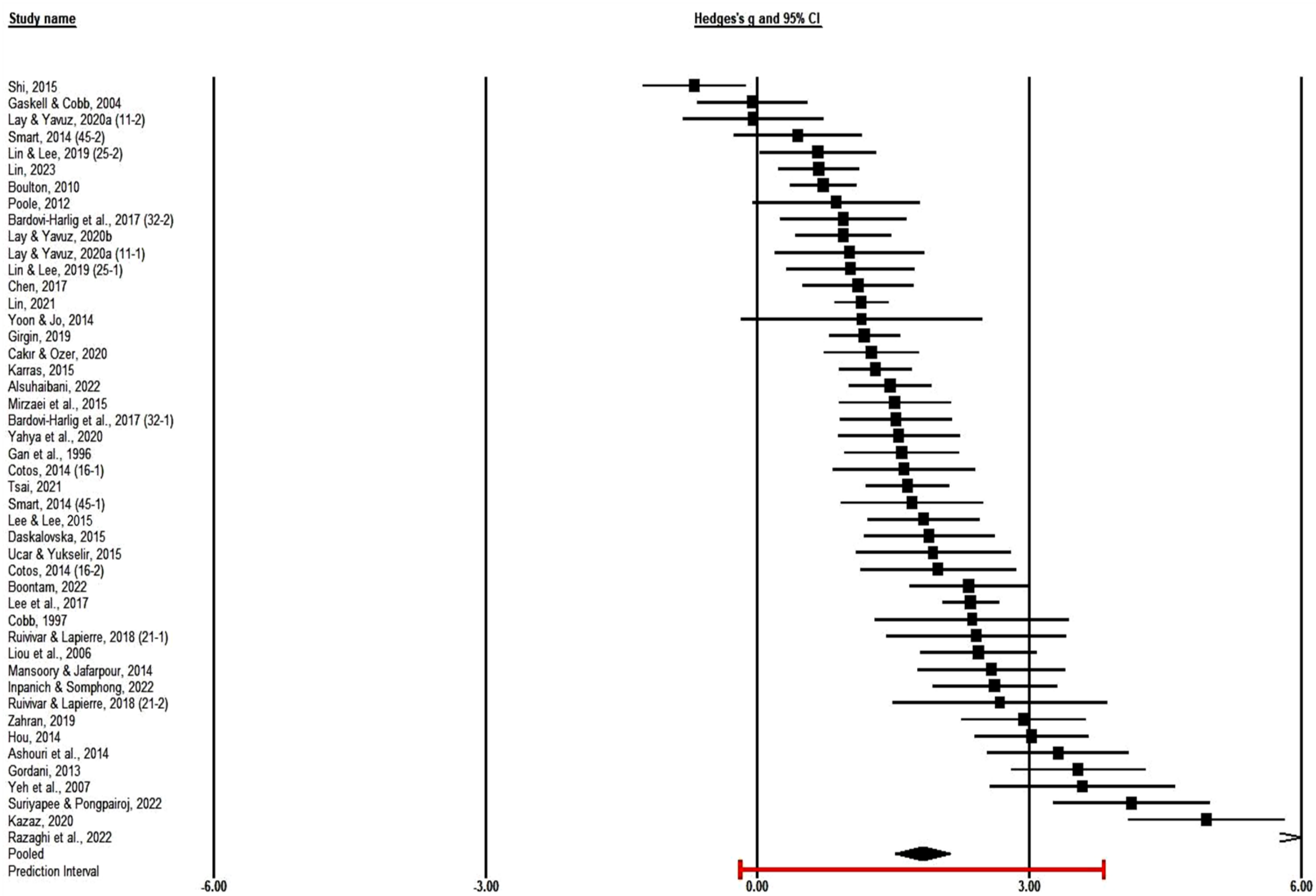

Fig. 2. The forest plot of effect sizes with 95% confidence intervals for 46 within-group (pre- and immediate post-tests) unique samples

The weighted (pooled) mean effect size is displayed by the black (closed) diamond at the bottom in Figure 2, which shows that there will be a 95% probability that the true mean effect size falls between 1.51 and 2.12. The capped red (gray) line below the diamond is the prediction interval showing there is a 95% chance that the effect size of a new study may fall between -0.19 and 3.82.

The weighted mean effect size calculated with all the 13 delayed-post-test within-group unique samples is 1.33, which could be considered ‘large’. This weighted mean effect size is presented at the bottom in Table 4 below, along with the effect sizes of each unique sample, which are sorted low to high.

Table 4. The weighted mean effect size and the effect sizes of 13 delayed-post-test within-group unique samples

ID	Study Name (Unique Sample ID)	Hedges's *g*	*SE*	*p*-value
45-2	Smart, 2014 (45-2)	0.17	0.36	0.64
25-2	Lin & Lee, 2019 (25-2)	0.67	0.33	0.04
12	Boulton, 2010	0.73	0.18	0.00
27	Lay & Yavuz, 2020b	0.95	0.27	0.00
25-1	Lin & Lee, 2019 (25-1)	1.03	0.36	0.00
15	Lin, 2021	1.15	0.16	0.00
45-1	Smart, 2014 (45-1)	1.16	0.37	0.00
36	Girgin, 2019	1.18	0.20	0.00
40	Lee & Lee, 2015	1.56	0.30	0.00
10	Daskalovska, 2015	1.89	0.37	0.00
41	Inpanich & Somphong, 2022	2.62	0.35	0.00
21-1	Ruivivar & Lapierre, 2018 (21-1)	2.70	0.53	0.00
21-2	Ruivivar & Lapierre, 2018 (21-2)	2.71	0.61	0.00
	Weighted Mean Effect Size	1.33	0.17	0.00
	Mean of Prediction Interval	1.33		

The Effect Sizes of Subgroups in Seven Moderator Variables

The number of effect sizes in a lot of subgroups under moderator variables is less than seven, so the comparison between subgroups is not very meaningful (see Borenstein et al., 2010). In order to circumvent this issue, the present meta-analysis just takes five as the threshold for analyzing the effects of moderators, not seven. Still, this research has found only eight statistically significant contrasts in some moderator variables for between-group or within-group analysis.

For the subgroup analysis of the third moderator variable (instructional status/age) in between-group contrasts, the *Q*-value is statistically significant: $p = .08$ ($p < .10$). There exist only two subgroups in this moderator variable for the between-group contrasts: Secondary (6 effect sizes) 1.60 (Hedges's *g*) and Undergraduate (28) 1.01.

For the subgroup analysis of the third moderator variable (instructional status/age) in within-group contrasts, the *Q*-value is statistically significant: $p = .03$ ($p < .10$). Actually, there is some seemingly difference between the effect sizes of subgroups: Secondary (5 effect sizes) 3.10 (Hedges's *g*); Undergraduate (35) 1.59; Graduate (2) 1.80; and Mixed (3) 1.56. The only statistically significant difference is, however, the one between Secondary and Undergraduate ($p < .01$).

For the subgroup analysis of the fourth moderator variable (region: categorized) in between-group contrasts, the *Q*-value is statistically significant: $p = .01$ ($p < .10$). Although there is some seemingly difference across the effect sizes of subgroups—East Asia (9 effect sizes) 0.67 (Hedges's *g*); Turkey (6) 1.15; Iran/Arab (12) 1.63; and ESL/Europe (7) 0.76—only two subgroup comparisons are statistically significant among the six (= 4 x 3 / 2) combinations of these subgroups: East Asia 0.67 vs. Iran/Arab 1.63; ESL/Europe 0.76 vs. Iran/Arab 1.63. The difference between the two effect sizes is huge, and the statistical significance of *Q*-value is robust with $p < .01$ and $p = .02$ for the two subgroup comparisons, respectively.

For the subgroup analysis of the fourth moderator variable (region: categorized) in within-group contrasts, the *Q*-value between the four subgroups is statistically significant ($p < .01$). However, only three comparisons are statistically significant among the six (= 4 x 3 / 2) combinations of these subgroups: East Asia 1.75, ESL/Europe 1.35, or Turkey 1.05 vs. Iran/Arab 2.76. The difference between the two effect sizes is huge, and the statistical significance of *Q*-value is robust with $p = .02$, $p < .01$, and $p < .01$ for the three subgroup comparisons, respectively.

For the subgroup analysis of the sixth moderator variable (type of corpus) in within-group contrasts, the *Q*-value in total is statistically significant: $p = .08$ ($p < .10$). However, only one comparison has at least five effect sizes in each of its subgroups: General (23) vs. Specialized (14). In addition, the difference between these two effect sizes (1.71 and 1.39, respectively) is not statistically significant: $p = .33$ ($p > .10$).

For the subgroup analysis of the seventh moderator variable (intervention duration) in between-group contrasts, the *Q*-value in total is statistically significant: $p = .07$ ($p < .10$). However, only one comparison is statistically significant among the combinations of these four subgroups: 'up to 1 week' 0.55 vs. '1.1–5.0 weeks' 1.34. The difference between the two effect sizes is huge, and the statistical significance of *Q*-value is robust with $p = .02$.

Discussion

The Overall Effect Size of Between-Group Unique Samples

As mentioned in the chapter Results, the weighted mean effect size (Hedges's *g*) 1.11 of between-group (immediate post-tests) contrasts is considered 'large' in terms of Plonsky and Oswald's (2014) benchmarks. Also, it is much larger than that (0.69 for Cohen's *d*) of overall L2 research and that (0.35 for Cohen's *d*) of educational technology in general (Plonsky & Oswald, 2014). In addition, it is much larger than that (0.51) of overall CALL (Plonsky & Ziegler, 2016). This implies that the corpus-based DDL approach could be one of the most effective and efficient instructional methods among others in EFL/ESL settings.

More importantly, the effect size (1.11) of this meta-analysis is in line with that (0.95 for Hedges's *g*) of the previous comprehensive L2-/DDL-related meta-analysis (Boulton & Cobb, 2017) where the between-group effect sizes of three unique samples are winsorized (reduced) to 3.00. Relative to the previous meta-analysis focusing only on L2 vocabulary in DDL (Lee et al., 2019), the between-group weighted mean effect size of the subgroup Vocabulary (including 22 effect sizes in it) in the current study is 1.20 (greater than 'large'), which is larger than the medium-sized one (0.74 for Hedges's *g*) of this previous vocabulary-related meta-analysis.

This indicates that the effectiveness of DDL has not been decreasing at all although this relatively new field has been matured over time. Regarding the maturity and effectiveness of a new academic discipline, Plonsky and Oswald (2014) argued as follows:

> More mature domains are therefore more likely to be examining relationships qualified by a more specific situation or criterion. In domains where this scenario is observed, theoretical maturity could be inversely correlated with outcomes, and a decrease in effect sizes would be obtained over time. (p. 895)

These two contradicting phenomena between DDL and other disciplines investigated by Plonsky and Oswald (2014) could be interpreted in two ways. First, the academic subdiscipline, DDL, is not matured yet since it has not been researched sufficiently, which will be discussed more in the chapter Conclusion. Second, DDL could be highly effective and might be more effective in the near future as well although there is observed a decreasing trend of instructional effectiveness in general.

Compared to the 34 between-group unique samples for immediate post-tests, the number of between-group unique samples for delayed post-tests is only 12. However, a statistically significant result is observed ($p < .01$) as well. The weighted mean effect size of delayed post-tests is between medium (0.70) and large (1.00), which indicates that the learning outcomes remain highly positive (0.83) in the long run as well as short-term (1.11) although it decreases over time. The 95% confidence interval of this mean effect size ranges rather widely from 0.40 ('small') to 1.26 (greater than 'large'), but the true mean effect size would be still positive anyways with a statistical significance ($p < .01$).

Relative to the previous meta-analysis focusing only on L2 vocabulary in DDL (Lee et al., 2019), the present study's effect size (0.83) for delayed post-tests (between-group) is approximately in line with that (0.64 for Hedges's *g*) of this previous meta-analysis. In addition, the current study's effect size (0.77) of the subgroup Vocabulary (between-group; delayed post-tests) is also in line with that (0.64) of this previous vocabulary-related meta-analysis.

The Overall Effect Size of Within-Group Unique Samples

The weighted mean effect size (Hedges's *g*) 1.81 of within-group contrasts is in line with that (1.50 for Hedges's *g*) of the previous comprehensive L2-/DDL-related meta-analysis (Boulton & Cobb, 2017) where the within-group effect sizes of 10 unique samples are winsorized (reduced) to 3.00. This within-group result is in line with the between-group one mentioned above and also implies that DDL could be one of the most effective instructional methods in EFL/ESL settings. Regarding the previous meta-analysis focusing only on L2 vocabulary in DDL, Lee et al. (2019) do not conduct a within-group analysis, so there is no comparable data in this previous study.

Compared to the weighted mean effect size (1.81) between the pre- and immediate post-tests, the one between the pre- and delayed post-tests still remains high (1.33), which is estimated with 13 unique samples. The number of unique samples is not sufficient relative to 46 (for the pre- and immediate post-test contrasts), but a statistically significant result is observed as well ($p < .01$). The effect size for delayed post-tests (within-group) is close to large (1.40), which implies that the learning outcomes remain highly positive (1.33) in the long term as well as short term (1.81) although it decreases over time. More importantly, the current study's effect size (1.33) for delayed post-tests (within-group) is in line with that (1.36 for Hedges's *g*) of the previous comprehensive meta-analysis (Boulton & Cobb, 2017).

The Effect Sizes of Subgroups in Seven Moderator Variables

The subgroup Secondary in the third moderator variable 'instructional status/age' (between-group) shows a much larger effect size (1.60) than that (1.01) of the subgroup Undergraduate. This difference is in line with Boulton and Cobb's (2017) results (Hedges's *g*): School (1.41) vs. University 1 and University 2–3 (0.96 and 0.45, respectively). Since the effect sizes are only six in the subgroup Secondary (between-group), however, a further interpretation might not be so meaningful as well as Boulton and Cobb (2017) do not provide a plausible reason for this phenomenon due to the same problem (only six samples in the subgroup School).

The subgroup Secondary in the third moderator variable 'instructional status/age' (within-group) shows a much larger effect size (3.08) than that (1.59) of the subgroup Undergraduate with a statistical significance of $p < .01$. This difference is in line with that of the between-group analysis above, but it is much larger than that of Boulton and Cobb's (2017) within-group results (Hedges's *g*): School (1.56) vs. University 1 and University 2–3 (1.41 and 1.27, respectively).

The subgroup Iran/Arab in the fourth moderator variable 'region' (between-group) shows a much larger effect size (1.64) than that (0.67) of the subgroup East Asia. This difference is in line with Boulton and Cobb's (2017) results (Hedges's *g*): Middle East (1.39) vs. Asia (0.84). This implies that learners' L1 and cultural backgrounds might have an influence on the effects of corpus-based DDL for EFL learners.

The subgroup Iran/Arab in the fourth moderator variable 'region' (between-group) shows a much larger effect size (1.64) than that (0.76) of the subgroup ESL/Europe with a statistical significance of $p = .02$. This difference is in line with Boulton and Cobb's (2017) results (Hedges's *g*): Middle East (1.39) vs. Europe (0.95) and North America (0.31). This huge difference indicates that DDL could be more effective/efficient in EFL settings than in ESL ones as Boulton and Cobb (2017) interpreted this interesting phenomenon as follows:

> Conversely, it was in Europe and North America that effect sizes were rather lower (though still reasonably robust), two regions where inductive, problem-solving approaches would seem more in line with prevailing cultures. One obvious possibility is that DDL was not different enough from traditional teaching in these parts of the world, and this was somewhat borne out by C/E [control vs. experimental] designs producing the lowest effect sizes. (p. 374)

The subgroup Iran/Arab in the fourth moderator variable 'region' (within-group) shows a much larger effect size (2.83) than that (1.75) of the subgroup East Asia. This difference is in line with Boulton and Cobb's (2017) results (Hedges's *g*): Middle East (2.07) vs. Asia (1.55). This phenomenon also fits well with the comparable results of the between-group analysis mentioned above.

The subgroup Iran/Arab in the fourth moderator variable 'region' (within-group) shows a much larger effect size (2.80) than that (1.36) of the subgroup ESL/Europe with a statistical significance of $p < .01$. This difference fits well with Boulton and Cobb's (2017) results (Hedges's *g*): Middle East (2.07) vs. Europe (1.15) and North America (0.95). In addition, this phenomenon is in line with the comparable results of the between-group analysis in this study.

The subgroup Iran/Arab in the fourth moderator variable 'region' (within-group) shows a much larger effect size (2.80) than that (1.05) of the subgroup Turkey with a statistical significance of $p < .01$. This provides another evidence on the strong effect of DDL in the subgroup Iran/Arab.

The subgroup '1.1–5.0 weeks' in the seventh moderator variable 'intervention duration' (between-group) shows a much larger effect size (1.34) than that (0.55) of the subgroup 'up to 1 week'. However, this difference is not in line with that of Boulton and Cobb's (2017) results: Medium (3 to 8 classes; 1 class = about 2 hours) (0.85 for Hedges's *g*) vs. Short (2 hours or less) (0.89). This means that there should be more investigation on this moderator variable, especially when it is considered that the longer instructional periods show less effect sizes in the present study: '5.1-11.9 weeks' (1.38) vs. '12 weeks or more' (0.85).

Conclusion

The above-mentioned problem—the shortage of primary quantitative studies in a large number of subgroups for corpus-based DDL—is also found in MALL, another relatively new field in ISLA. In Burston and Giannakou's (2022) meta-analysis on MALL in general, only one moderator variable shows a statistically significant difference between its subgroups. Also, there are only three subgroup comparisons that are statistically significant in Yoon and Lee's (2023) meta-analysis on MALL (Mobile-Assisted Language Learning) combined with feedback.

Consequently, as corpus linguists and researchers (e.g., Hunston, 2022; Lee, 2007; O'Keeffe et al., 2007) argue, it is recommended that there should be more diverse primary experimental studies in new academic sub-disciplines such as corpus-based DDL approach in order to help evaluate this relatively new field in ISLA more precisely and meticulously. For instance, more studies are needed on other language skills/aspects than vocabulary, grammar, and writing in DDL. Also, it would be helpful if there were more research on learners with a low L2 proficiency or in secondary schools and on learner corpora utilized together with native speakers' corpora (e.g., Granger, 1998; Lee, 2007).

Only with more meticulous/precise academic assessment of each subgroup in DDL as mentioned above, both teachers and students can have a stronger confidence in implementing a specific DDL pedagogy of their choice in their classrooms. Since there are highly diverse DDL approaches—which may have led to the widespread individual effect sizes revealed in this meta-analysis—the very large weighted mean effect size of overall DDL itself might not be sufficient to persuade real-world teachers and students to adopt such a relatively unfamiliar and time-consuming pedagogy as DDL.

Moreover, there should be more higher-quality experimental studies on DDL. Through the search process in ERIC, 73 primary quantitative studies are found on corpus-based DDL in EFL or ESL. However, only 46 among these 73 studies have proper research designs and report sufficient statistical data with which effect sizes can be calculated.

In addition, some primary studies do not fully report their descriptive statistics: the mean, standard deviation, and sample size that are invaluable data inputs for other colleagues to investigate/interpret the research results more intensively and extensively. Instead of this kind of necessary descriptive statistics, some primary research articles provide partially with the inferential statistics such as the *F*-statistic or only the finally-estimated effect sizes themselves.

More importantly, only 34 out of 54 unique samples are utilized for between-group contrasts, which means that 20 unique samples do not have a comparison group. Although the within-group analysis itself can investigate the effectiveness of an instructional method, the efficiency is more important in ISLA. Contrary to such academic disciplines as medicine, there is actually no case in the real-world classrooms to compare an instructional treatment with no treatment at all such as placebos in medical experiments. In ISLA, only the level of efficiency that is assessed by the between-group analysis can persuasively guide both the teachers and students to choose the best possible instructional option for their limited resources and class hours in the real world.

Last but not least, there are only 12 between-group and 13 within-group unique samples out of 54 in total that are utilized for a delayed post-test, and only two studies (four unique samples) conduct post-tests that are delayed for more than four weeks. This is quite troublesome in the educational perspective, because the learning outcomes should remain positive in the long run as well as short-term.

References

Alsuhaibani, Z. (2022). Developing EFL students' pragmatic competence: The case of compliment responses. Language Teaching Research, 26, 847–866.
Ashouri, S., Arjmandi, M., & Rahimi, R. (2014). The impact of corpus-based collocation instruction on Iranian EFL learners' collocation learning. Universal Journal of Educational Research, 2, 470–479.
Barabadi, E., & Khajavi, Y. (2017). The effect of data-driven approach to teaching vocabulary on Iranian students' learning of English vocabulary. Cogent Education, 4, 1–13.
Bardovi-Harlig, K., Mossman, S., & Su, Y. (2017). The effect of corpus-based instruction on pragmatic routines. Language Learning & Technology, 21, 76–103.
Biber, D. (1988). Variation across Speech and Writing. Cambridge University Press.
Boontam, P. (2022). The effect of teaching English synonyms through data-driven learning (DDL) on Thai EFL students' vocabulary learning. Shanlax International Journal of Education, 10, 80–91.
Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to Meta-analysis. Wiley.
Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. Research Synthesis Methods, 1, 97–111.
Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2022). Comprehensive Meta-Analysis (Version 4) [Computer software]. Biostat.
Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. Language Learning, 60, 534–572.
Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. Language Learning, 67, 348–393.
Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. Language Learning and Technology, 25, 66–89.
Burston, J., & Giannakou, K. (2022). MALL language learning outcomes: A comprehensive meta-analysis 1994–2019. ReCALL, 34, 147–168.
Chen, L. (2017). Corpus-aided business English collocation pedagogy: An empirical study in Chinese EFL learners. English Language Teaching, 10, 181–197.
Chen, M., & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. International Journal of Corpus Linguistics, 23, 335–369.
Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? System, 25, 301–315.
Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. Computer Assisted Language Learning, 12, 345–360.
Cobb, T., & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.), Cambridge Handbook of English Corpus Linguistics (pp. 478–497). Cambridge University Press.
Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Lawrence Erlbaum Associates.
Cotos, E. (2014). Enhancing writing pedagogy with learner corpus data. ReCALL, 26, 202–224.
Çakır, İ., & Özer, M. (2020). Fostering intuitive competence in L2 for a better performance in EAP writing through fraze.it in a Turkish context. Education and Information Technologies, 25, 5405–5426.
Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. Computer Assisted Language Learning, 28, 130–144.
Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations: A meta-analysis. International Journal of Corpus Linguistics, 19, 443–477.

Francis, G. (1993). A corpus-driven approach to grammar: Principles, methods and examples. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), Text and Technology: In Honour of John Sinclair (pp. 137–156). John Benjamins.

Gan, S.-L., Low, F., & Yaakub, N. F. (1996). Modeling teaching with a computer-based concordancer in a TESL preservice teacher education program. Journal of Computing in Teacher Education, 12, 28–32.

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? System, 32, 301–319.

Girgin, U. (2019). The effectiveness of using corpus-based activities on the learning of some phrasal-prepositional verbs. TOJET: The Turkish Online Journal of Educational Technology, 18, 118–125.

Gordani, Y. (2013). The effect of the integration of corpora in reading comprehension classrooms on English as a foreign language learners' vocabulary development. Computer Assisted Language Learning, 26, 430–445.

Granger, S. (Ed.). (1998). Learner English on Computer. Longman.

Hedges, L., & Olkin, I. (1985). Statistical Methods for Meta-analysis. Academic Press.

Hoey, M. (2005). Lexical Priming: A New Theory of Words and Language. Routledge.

Hou, H.-I. (2014). Teaching specialized vocabulary by integrating a corpus-based approach: Implications for ESP course design at the university level. English Language Teaching, 7, 26–37.

Hunston, S. (2022). Corpora in Applied Linguistics (2nd ed.). Cambridge University Press.

Hunston, S., & Francis, G. (2000). Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English. John Benjamins.

Hwang, S.-D. (2020). Meta-analysis Using R (2nd ed.). Hakjisa.

Inpanich, P., & Somphong, M. (2022). The effects of the integration of indirect feedback and concordances on improving grammatical accuracy in Thai EFL students' writing. English Language Teaching, 15, 25–38.

Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. English Language Research Journal, 4, 27–45.

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. English Language Research Journal, 4, 1–16.

Karras, J. N. (2015). The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam. ReCALL, 28, 166–186.

Kartal, G., & Yangineksi, G. (2018). The effects of using corpus tools on EFL student teachers' learning and production of verb-noun collocations. PASAA: Journal of Language Teaching and Learning in Thailand, 55, 100–125.

Kazaz, İ. (2020). Alternative vocabulary assessment: Using concordance line activities for testing lexical knowledge. International Online Journal of Education and Teaching, 7, 1221–1237.

Kilimci, A. (2017). Integrating cognitive linguistics insights into data-driven learning: Teaching vertical prepositions. Journal of Language and Linguistic Studies, 13, 681–719.

Lay, K. J., & Yavuz, M. A. (2020a). Data-driven learning of academic lexical bundles below the C1 level. Language Learning & Technology, 24, 176–193.

Lay, K. J., & Yavuz, M. A. (2020b). Targeting Turkish-to-English interlingual interference through context-heavy data-driven learning. SAGE Open, 10, 1–12.

Lee, D. J. (2007). Corpora and the Classroom: A Computer-aided Error Analysis of Korean Students' Writing and the Design and Evaluation of Data-driven Learning Materials [Unpublished doctoral dissertation]. University of Essex, United Kingdom

Lee, H., & Lee, J. H. (2015). The effects of electronic glossing types on foreign language vocabulary learning: Different types of format and glossary information. Asia-Pacific Education Researcher, 24, 591–601.

Lee, H., Warschauer, M., & Lee, J. H. (2017). The effects of concordance-based electronic glosses on L2 vocabulary learning. Language Learning & Technology, 21, 32–51.

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. Applied Linguistics, 40, 721–753.

Lenhard, W., & Lenhard, A. (2016). Computation of Effect Sizes. Psychometrica. https://www.psychometrica.de/effect_size.html

Lin, M. H. (2021). Effects of data-driven learning on college students of different grammar proficiencies: A preliminary empirical assessment in EFL classes. SAGE Open, 11, 1–15.

Lin, M. H., & Lee, J.-Y. (2019). Pedagogical suitability of data-driven learning in EFL grammar classes: A case study of Taiwanese students. Language Teaching Research, 23, 541–561.

Lin, Y.-Y. (2023). Teaching grammar and vocabulary for the TOEIC test with corpora: The case of lower intermediate learners. Taiwan Journal of TESOL, 20, 67–113.

Liou, H.-C., Chang, J. S., Chen, H.-J., Lin, C.-C., Liaw, M.-L., Gao, Z.-M., Jang, J.-S. R., Yeh, Y., Chuang, T. C., & You, G.-N. (2006). Corpora processing and computational scaffolding for a web-based English learning environment: The CANDLE project. CALICO Journal, 24, 77–95.

Lipsey, M., & Wilson, D. (2001). Practical Meta-analysis. SAGE Publications.

Mansoory, N., & Jafarpour, M. (2014). Teaching semantic prosody of English verbs through the DDL approach and its effect on learners' vocabulary choice appropriateness in a Persian EFL context. Advances in Language and Literary Studies, 5, 149–161.

Mirzaei, A., Domakani, M. R., & Rahimi, S. (2015). Computerized lexis-based instruction in EFL classrooms: Using multi-purpose LexisBOARD to teach L2 vocabulary. ReCALL, 28, 22–43.

Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. English Corpus Studies, 22, 1–18.

Moon, S., & Oh, S.-Y. (2017). Unlearning overgenerated "be" through data-driven learning in the secondary EFL classroom. ReCALL, 30, 48–67.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. Language Learning, 50, 417–528.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). From Corpus to Classroom: Language Use and Language Teaching. Cambridge University Press.

Paquot, M., & Plonsky, L. (2017). Quantitative research methods and study quality in learner corpus research. International Journal of Learner Corpus Research, 3, 61–94.

Plonsky, L., & Brown, D. (2015). Domain definition and search techniques in meta-analyses of L2 research (Or why 18 meta-analyses of feedback have different results). Second Language Research, 31, 267–278.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. Language Learning, 64, 878–912.

Plonsky, L., & Ziegler, N. (2016). The CALL-SLA interface: Insights from a second-order synthesis. Language Learning & Technology, 20, 17–37.

Poole, R. (2012). Concordance-based glosses for academic vocabulary acquisition. CALICO Journal, 29, 679–693.

Razaghi, A., Faruji, L. F., & Salehi, M. (2022). Audio podcast retelling versus corpus-based learning and vocabulary knowledge development of English language learners. Anatolian Journal of Education, 7, 181–196.

Rezaee, A. A., Marefat, H., & Saeedakhtar, A. (2015). Symmetrical and asymmetrical scaffolding of L2 collocations in the context of concordancing. Computer Assisted Language Learning, 28, 532–549.
Ruivivar, J., & Lapierre, C. (2018). Learning outcomes and learners' impressions of parallel and monolingual concordancers. In P. Taalas, J. Jalkanen, L. Bradley & S. Thouësny (Eds.), Future-proof CALL: Language Learning as Exploration and Encounters - Short Papers from EUROCALL 2018 (pp. 278–283). Research-publishing.net.
Satake, Y. (2022). The effects of corpus use on L2 collocation learning. The JALT CALL Journal, 18, 34–53.
Shi, J. (2015). An analysis of the application of wikipedia corpus on the lexical learning in the second language acquisition. English Language Teaching, 8, 171–180.
Sinclair, J. (Ed.). (1987). Collins Cobuild English Language Dictionary. HarperCollins.
Smart, J. (2014). The role of guided induction in paper-based data-driven learning. ReCALL, 26, 184–201.
Sripicharn, P. (2003). Evaluating classroom concordancing: The use of concordance-based materials by a group of Thai students. Thammasat Review, 8, 203–236.
Stevens, V. (1991). Concordance-based vocabulary exercises: A viable alternative to gap-fillers. English Language Research Journal, 4, 47–61.
Suriyapee, P., & Pongpairoj, N. (2022). Corpus to enhance verbal complements among low English proficiency Thai learners. PASAA: Journal of Language Teaching and Learning in Thailand, 63, 205–231.
Tono, Y. (2002). The Role of Learner Corpora in SLA Research and Foreign Language Teaching: The Multiple Comparison Approach [Unpublished doctoral dissertation]. University of Lancaster, United Kingdom.
Tsai, Y.-R. (2021). Exploring the effects of corpus-based business English writing instruction on EFL learners' writing proficiency and perception. Journal of Computing in Higher Education, 33, 475–498.
Uçar, S., & Yükselir, C. (2015). The effect of corpus-based activities on verb-noun collocations in EFL classes. TOJET: The Turkish Online Journal of Educational Technology, 14, 195–205.
Yahya, N., Alotaibi, H., & El-Dakhs, D. A. S. (2020). Parallel corpora in EFL writing classrooms: Are they effective? International Journal of Computer-Assisted Language Learning and Teaching, 10, 23–39.
Yeh, Y., Liou, H.-C., & Li, Y.-H. (2007). Online synonym materials and concordancing for EFL college writing. Computer Assisted Language Learning, 20, 131–152.
Yoon, H., & Jo, J. W. (2014). Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. Language Learning & Technology, 18, 22–96.
Yoon, K.-H., & Lee, D. J. (2023). The cognitive effects of MALL combined with feedback in EFL/ESL: A meta-analysis. Journal of the Korea English Education Society, 22, 97–123.
Zahran, F. A. (2019). The impact of corpus on EFL pre-service teachers self-directed learning and oral proficiency. Language Teaching Research Quarterly, 13, 85–105.

Authors' Information

Yoon, Kyung-Ho: Gonghang Middle School, Teacher, 1st Author https://orcid.org/0009-0000-5665-7165

Lee, Dong Ju: Korea National University of Education, Professor, Corresponding Author https://orcid.org/0000-0002-9906-835X