

Feedback in Computer-Assisted Language Learning: A Meta-Analysis

August 2020 – Volume 24, Number 2

Adnan F. Saad Mohamed

Washington State University

<adnan.mohamed@wsu.edu; adnanfsmohamed@gmail.com>

Abstract

Feedback is a well-known advantage for language learning. Primary studies on feedback in computer-assisted language learning (CALL) demonstrates that feedback has a significant effect on student language learning. Previous reviews (e.g., Azevedo & Bernard, 1995; Kang & Han 2015; Li, 2010) provided important insights on language learning. However, these reviews showed that there has never been a meta-analysis synthesizing the effectiveness of feedback in CALL studies and the moderators moderating the effect of feedback in CALL. With the aim of summarizing years of research on feedback in CALL studies and identifying the moderators of feedback in CALL, a meta-analysis was conducted. By establishing rigorous inclusion and exclusion criteria, the investigator located 21 primary studies that met the inclusion and exclusion criteria. The findings indicated under the Random-Effects (RE) model that feedback in CALL has a significant medium effect size on student language learning outcomes ($g = 0.56$). The results also showed that the effect of feedback is moderated by a host of variables, including learners' mother tongue, intervention provider (i.e., teacher, researcher), target language, and so on. The study concluded that feedback in CALL is a promising field for language learning and provided implications for teachers and future research.

Keywords: feedback, computer-assisted language learning

Feedback is a piece of post-commentary information given on learners' language performance by a teacher, program, or application (Duhon, House, Hastings, Poncy, & Solomon, 2015; Liutkus, 2012). Research on language learning views feedback as essential for learners at all levels from kindergarten through 12th grade (K–12) to college courses (Fajfar, Campitelli, & Labollita, 2012). The importance of feedback lies in its role in triggering learners' attention to notice the discrepancies between their output and the desired performance (Kang & Han, 2015). In search of help for teachers providing feedback for every learner, a number of researchers in language learning suggested CALL. CALL is an interactive use of technology (Beechler, & Williams, 2012; Rashidi, & Babaie, 2013). It facilitates language learning by providing opportunities (Stickler & Shi, 2016) to practice the language in interactive environments in which learners can learn through using

videos, pictures, audios, games, social media, applications, and online discussions with native speakers (Sydorenko, 2010). CALL has gained popularity for providing feedback in a quick manner. Many studies (e.g., Dongyu, Fanyu, & Wanyi, 2013; Fahim, & Haghani, 2012) have shown that feedback in CALL had a significant contribution to students' language learning outcomes.

Sociocultural theorists acknowledged the benefits of feedback on language learning (e.g., Ai, 2017; Anh & Marginson, 2013; Dongyu, et al., 2013; Elola & Oskoz, 2016; Hung, 2016). In their opinion, language learning should not be isolated to where learners work unassisted and unmediated. "Sociocultural theorists emphasize the importance of the provision of co-participation and guided practice as a prominent feature in activity settings where expertise is distributed, practiced, and shaped in order to produce a common product or artifact" (Englert, Mariage, & Dunsmore, 2006, p. 209).

Sociocultural theorists consider feedback to be an essential component of language learning informing learners about their performance. The main purpose of feedback is to scaffold learners at their current level to approach the desired level (Ilies, Judge, & Wagner, 2010; Meo, 2013). Scaffolding can be a dictionary, application, or feedback (Dongyu et al., 2013; Elola & Oskoz, 2016). Scaffolding is also interpreted as support from a knowledgeable person or application to the less knowledgeable person to accomplish a task which s/he could not do by her/himself (Kargar & Tayebipour, 2015). Scaffolding is often considered as an advantage to strengthen learning outcomes (Masuda, Arnett, & Labarca, 2015).

Since the role of feedback in learning is theoretically acknowledged, a growing body of traditional reviews and empirical research has built a case for the benefits of feedback for students' language learning. One of the highly cited traditional reviews is Hattie and Timperley's (2007) work. In their review, Hattie and Timperley (2007) drew conclusions from meta-analyses on feedback. Their findings suggested that feedback on the task and learning process is more beneficial to students' learning outcomes than the feedback given at learners' characteristics.

Recent meta-analyses on the effects of feedback on language showed that feedback has a positive impact on language learning. For example, Li's (2010) meta-analysis included 22 published studies and 11 unpublished ones on the effect of corrective feedback on second language learning between 1968 and 2007. The findings showed a positive effect of the feedback in fixed-effects (EF) $d= 0.61$ and random-effects (RE) model $d= 0.64$. Li's (2010) analysis also included moderators of the feedback effect. The moderator analysis showed the following. First, unpublished studies had a greater effect size than the published ones. Second, lab-based studies' effect was higher than classroom-based studies. Third, studies with a short-term intervention had a larger effect size than those with a long-term treatment. Fourth, implicit feedback effect was greater than explicit feedback. Studies conducted in foreign language settings generated a larger effect size than those in second language settings. However, this meta-analysis had only six studies on feedback in CALL. It remains questionable whether the findings apply to feedback in CALL.

Kang and Han (2015) conducted a meta-analytic approach on 21 studies published between 1980 and 2013 about whether written corrective feedback helps to improve the grammatical accuracy of second language writing and what factors mitigate the efficacy of written corrective feedback. The outcome of the study suggested that written corrective feedback had a moderate effect ($d= 0.68$) on the grammatical accuracy of second language writing. In addition, Kang and Han (2015) conducted a moderator analysis. Learners' language proficiency, treatment sessions, target language, time of publication, and writing genres were significant moderators of written corrective feedback effects. Feedback in a second language setting had a larger effect size than those in foreign language contexts. However, the findings of Kang and Han (2015) might not be extrapolated to

feedback in CALL because one of their exclusion criteria is to exclude any feedback given by a computer.

Van der Kleij, Feskens, and Eggen (2015) conducted a meta-analysis of 40 studies to investigate the effectiveness of feedback in computer instruction environments on student learning outcomes. This meta-analysis included 30 published studies and 10 dissertations published between 1960 and 2012. The results indicated the elaborative feedback has a larger effect size (0.49) than knowledge of correct results (0.05) and knowledge of response (0.32). Furthermore, the study suggested that immediate feedback is more effective for student lower-level outcomes than delayed feedback. However, the authors did not conduct a moderator analysis.

Previous meta-analyses about feedback have different foci, perspectives, and results reported on a different set of primary studies (e.g., Kang & Han, 2015; Li, 2010; Van der Kleij et al., 2015). Kang and Han (2010) did not include studies where feedback is given by a computer. On the contrary, Li (2010) included a few studies ($k=6$) on feedback where feedback is delivered by computer. This small number makes it questionable whether the findings apply to feedback in CALL.

Although Kang and Han (2010) and Li (2010) investigated the moderators of the feedback effect, they had different results. For example, Li (2010) reported that studies conducted in foreign language settings generated a larger effect size than those in second language settings. However, Kang and Han (2010) indicated that second language settings' studies had a greater effect size than those in foreign language settings. Since Li (2010) did not show which context is greater with CALL studies, this makes it important to identify which context is greater with studies in CALL feedback research.

To date, there is no meta-analysis that is devoted entirely to feedback in CALL. In previous meta-analyses, CALL was a part of the overall analysis. For example, Van der Kleij et al. (2015) meta-analysis was conducted on a computer-based learning environment. A computer-based learning environment refers to the use of technology to support learning (Wang, Wu, Kirschner, & Spector, 2018; Van Laere, Agirdag, & Van Braak, 2016). It includes multiple disciplines (e.g., psychology, literacy, math, computer science). Therefore, Van der Kleij et al. (2015) findings represent the effect of certain types (i.e., elaborative, knowledge of correct results, knowledge of response) and time (i.e., immediate, delayed) of feedback in the computer-based learning environment in different disciplines. Van der Kleij et al.'s (2015) meta-analysis included several studies in CALL. However, the authors did not conduct a moderator analysis.

Neither of the meta-analyses provided information regarding moderators such as intervention modeling and intervention provider (i.e., teacher, researcher). This necessitates the need to conduct a meta-analysis demonstrating factors impacting the effectiveness of feedback in CALL across studies. Although the effects of several factors (e.g., modeling, intervention provider) on feedback in CALL have not been explored by previous meta-analyses, this meta-analysis is not limited to these two factors. It also includes other moderators such as intervention length, measures of proficiency, mother tongue, publication type, research context, and target language. This effort might help understand feedback in CALL under various moderators.

There is a strong justification for this meta-analysis and a variation of findings in moderators impacting feedback in studies related to feedback. To date, it is questionable which moderators could influence feedback effect in CALL literature. Due to a lack of a meta-analysis focusing on the factors influencing the effects of CALL feedback on language learning, there is a need for an analysis to help understand under which factors moderate CALL feedback. Demonstrating which moderators are effective might help CALL feedback users to consider factors when providing

feedback to learners in CALL environments. Therefore, the findings of this research can be beneficial informing theory and practice to consider certain variables.

Research Questions

This brief review has shown there is a significant volume of research on the effectiveness of feedback on language learning. However, little is known about studies synthesizing the influence of specific moderators (e.g., intervention modeling, intervention provider) and the other aforementioned moderators on feedback effect in CALL research.

The examined moderators in this meta-analysis are considered independent variables. The CALL feedback effect sizes extracted from the included primary studies is a dependent variable. This meta-analysis attempts to answer:

1. What is the overall effect of CALL feedback on language learning?
2. What are the variables that moderate the effect of CALL feedback on language learning?

Method

To gain a greater understanding of the overall effect of feedback in CALL on language learning and its moderators, a meta-analysis was conducted. Meta-analysis is a statistical technique to combine the results from independent experimental studies (Borenstein, Hedges, Higgins, & Rothstein, 2005; Borenstein, Hedges, Higgins, & Rothstein 2009; Lipsey & Wilson, 2001). The investigator utilized meta-analysis procedures to collect and analyse studies. These procedures included: (a) comprehensive search to identify potential target studies; (b) coding study characteristics (moderator variables); and (c) statistical analysis to calculate the overall effect of feedback and identify the influences of the potential moderators.

Comprehensive Searching

From November 2018-April 2019, the investigator conducted a comprehensive and systematic search on educational electronic databases (e.g., Academia, ERIC, Google Scholar, ResearchGate, PsycINFO, ProQuest). The search was to identify articles, dissertations, and thesis about feedback in CALL. Variations of key terms (*feedback**, *computer-based feedback**, *computer-provided feedback**, *computer* and feedback**, or *feedback* and CALL**) were utilized. This search did not have any time restriction, and any study was published up to 2019 was considered. The electronic search located 2884 titles.

After removing the duplicates, the results led to 1,211 abstracts to be screened. In the abstract review, each abstract was read to determine whether it related to an article, dissertation, or Master's thesis. The abstract review resulted in 374 papers to be retrieved. For the full review, the investigator reviewed the 374 papers to decide whether the study met the criteria (see Table 1 for more information about inclusion/exclusion criteria). The full review led to 19 studies. In ancestral search, the investigator conducted an ancestral search including the reference of the 19 selected papers in the full a review and previous meta-analyses (Azevedo & Bernard, 1995; Kang & Han 2015; Li, 2010; Van der Kleij et al., 2015) on feedback in the computer learning environment. Prior to the ancestral search, 19 studies were selected. When the investigator conducted the ancestral search, two studies were added. In total, 21 studies were analysed for this study analysis.

Table 1. Inclusion/Exclusion Criteria.

Inclusion criteria	Exclusion criteria
1. Must be a CALL study including feedback as an independent variable.	1. Did not investigate the effect of feedback in on linguistic learning features.
2. Must report sufficient quantitative data such <i>M</i> , <i>SD</i> , and number of participants to calculate the effect size.	2. Did not provide the effect size or information to calculate effect size.
3. Must be in English.	3. It is not in English.
4. Must have participants where the target language differed from their mother tongue.	4. Did not show the difference between participant mother tongue and target language (e.g., Ifenthaler, 2010).
5. Must have feedback given by computer.	5. Did not include outcome measures in the investigation (e.g., Heift, 2010).
6. Must be experimental or quasi-experimental design.	6. Did not use experimental or quasi-experimental design.
7. Must have a treatment and control or a comparison group.	7. Did not include a control group or a group that can be considered a comparison group.

In summary, the comprehensive search resulted in 21 studies with a total of 1313 participants. They were published between 1992 and 2016. They included learners from undergraduate and graduate levels. Studies with learners from elementary, middle, and secondary schools did not meet the inclusion criteria. The studies included in the analysis are 12 articles, one Master's thesis, and 8 Ph.D. dissertations. All studies involved an immediate posttest.

Coding Study Characteristics

The investigator coded each article, dissertation, and thesis into categories. Because of the difficulty and significance of coding, meta-analyses (e.g., Adesope, Trevisan, & Sundararajan, 2017; Belland, Walker, Kim, & Leftler, 2017; Kang & Han, 2015; Li 2010; Van der Kleij et al., 2015) were consulted to establish a preliminary coding scheme for this work. The creation of a coding scheme was a continuous process of repeated modifications and revisions. The scheme contained 13 major variables: (a) educational level, (b) intervention length, (c) intervention modeling, (d) intervention provider, (e) intervention status, (f) language proficiency, (g) measures of proficiency, (h) mother tongue, (i) publication type, (j) research context, (l) research setting, (m) subject domain, and (n) target language. These moderators are described below.

First is the educational level. In this study, it includes three levels: Graduate, undergraduate, and mixed. Freshmen, sophomore, junior, and senior were categorized as an undergraduate. Masters or a doctoral level were identified as a graduate. If a study had undergraduates and graduates, their level was coded as mixed. For the intervention length, feedback literature (e.g., Li, 2010) cataloged the intervention length as short as 10 minutes or 2 hours long. It is significant to investigate the influence of this variable on the overall effect of feedback in CALL research. Based on Li's (2010)

study, if the length of the intervention was less than 50 minutes, it was coded as “short”; 60 – 120-minute intervention was considered “medium”; a long-time intervention exceeded 120 minutes.

With regard to the intervention modeling, it is surprising that modeling has not been investigated in feedback research. The intervention modeling is defined as whether participants had trained to use the application or program. It is important to explore if learners received modeling and the effect of modeling on the overall effect of feedback in CALL. To build a case for the effectiveness of modeling in feedback CALL research, this meta-analysis compares studies that indicated modeling was “modeled” and the studies that did not report it. Similar to intervention modeling, the intervention provider has not been examined. The intervention provider is an individual(s) delivering the treatment. The intervention provider is either an instructor(s) or researcher(s). Similar to intervention modeling, the intervention provider has not been investigated in feedback literature. In order to have a convincing case for the significance of the intervention provider, this study compares interventions given by instructors and interventions given by researchers.

Another moderator is the intervention status. It was classified as the intervention developed (i.e., an application created by the author or researcher) for the study or literature-based. Literature-based interventions refer to applications/programs used in other studies. Comparing developed interventions and literature-based ones and their impact on the overall effect of feedback in CALL can provide teaching and research implications for future research in CALL feedback literature. Sixth is the language proficiency. It is a very important moderator. Previous meta-analyses (e.g., Li, 2010) have shown that it is a statistically significant moderator. However, feedback literature lacks evidence regarding the significance of language proficiency in CALL feedback research. Due to its significance, it is included in this meta-analysis. In this study, Learner language proficiency was beginner, intermediate, advanced, or mixed if participants differed in their proficiency (Kang & Han, 2015; Li, 2010).

Similarly, the measures of proficiency are a significant moderator in the previous meta-analysis (e.g., Kang & Han, 2015; Li, 2010). Due to its significance, it is included in this meta-analysis. In this meta-analysis, measures of language proficiency were labeled as a pretest, class enrolment, or placement test. Eight is the learner mother tongue. The mother tongue is the learner’s first language. It is another understudied variable in feedback CALL literature. In this meta-analysis, the learner’s mother tongue is labeled as identified in the included studies. In studies with participants differed in their mother tongue, participant mother tongue was coded as mixed.

According to Li, (2010), publication type refers to whether the study is an article, thesis, or dissertation. It is important to include dissertations because of the “tendency on the part of researchers, reviewers, and editors to submit, accept, and publish studies that report statistically significant results” (Cornell & Mulrow 1999, p. 311). The tenth is the research setting. Previous meta-analyses (e.g., Kang & Han, 2015; Li, 2010) related to feedback in language research have shown different results across research settings. The research setting is categorized as either a foreign language (FL) or a second language (SL). The foreign language setting is about a language being learned where the language is a not mother tongue of the country (e.g., teaching English in China). The second language setting is the primary language of the community (e.g., Teaching English in the United Kingdom) (Kang & Han, 2015).

In this study, the research context is where the study took place. The study conducted either in a classroom or computer laboratory (CL). According to Li (2010), there is a difference in feedback between classroom studies and CL studies. Whereas the feedback in CL is about the intervention, the feedback in classrooms varies and might be difficult to measure. Therefore, the effect of feedback could differ between contexts. Twelfth is the subject domain. It refers to writing,

speaking, listening, reading, grammar, or pronunciation. It is not known the effect of the subject domain as a moderator on the overall effect of feedback in CALL. Therefore, it is an interest of this study to identify whether they moderate the overall effect of feedback in CALL. Last not the least is the target language. The target language is the language students are learning. The target language was categorized as reported in the included studies (Kang & Han, 2015). It is worthwhile to explore whether the target language can moderate the effect of feedback in CALL.

Data Analysis

The analysis was conducted by utilizing a meta-analysis software called comprehensive meta-analysis (CMA) (Version 2.2.064). CMA yielded the results of *Q*-tests, *I-square* statistics, plotting availability bias, FE models, and RE models (Borenstein et al., 2005; Li, 2010). The data analysis included reporting posttest effect size (see Table 2) and moderator analyses (see Tables 4, 5).

Table 2. Random-effects Model: Immediate Posttest.

Study name	Sample Size Total	Statistics for each study					
		Hedges' g	SE	Variance	Lower Limit	Upper Limit	Z-Value
Bationo (1992)	56	0.53	0.31	0.10	0.08	1.13	1.72
Gómez (2016)	71	1.75*	0.28	0.08	1.21	2.29	6.32
Hanson (2008)	28	0.05	0.39	0.15	0.72	0.82	0.12
Heift and Rimrott (2008)	28	1.35*	0.42	0.18	0.52	2.18	3.19
Hincks and Edlund (2009)	14	1.30*	0.56	0.31	0.25	2.39	2.33
Hsieh (2007)	52	1.32*	0.35	0.12	0.34	2.00	3.78
Kim (2009)	28	1.13*	0.45	0.20	-0.20	2.02	2.52
Kregar (2011)	87	0.78*	0.22	0.05	-0.61	1.21	3.51
Lavolette (2014)	112	0.17	0.19	0.04	-0.05	0.54	0.92
Lavolette et al (2015)	32	0.06	0.34	0.12	-0.24	0.74	0.18
Lee et al., (2012)	72	0.41	0.24	0.06	-0.23	0.88	1.74
Moreno (2007)	59	0.27	0.26	0.07	-0.05	0.79	1.04
Murphy (2007)	225	0.03	0.13	0.02	0.39	0.29	0.22
Nagata and Swisher (1995)	32	0.40*	0.18	0.03	-1.85	0.75	2.23
Nagata (1993)	34	0.27	0.34	0.11	-0.15	0.93	0.82
Neri and Strik (2008)	30	-1.10*	0.38	0.15	0.87	-0.35	-2.88
Park (1998)	63	0.44	0.30	0.09	-0.15	1.03	1.46
Rashidi and Babaie (2013)	120	1.28*	0.21	0.04	0.87	1.69	6.12
Sachs (2012)	78	0.11	0.24	0.06	-0.36	0.58	0.45
Sanz and Morgan-Short (2004)	69	0.95*	0.25	0.06	0.45	1.44	3.77

Sauro (2009)	23	0.80	0.44	0.19	0.06	1.66	1.82
Random		0.56*	0.13	0.02	0.31	0.82	4.40

Note. $*= p < .05$, K =number of studies, SE =standard Error, CI = confidence interval, Df = degree of freedom, g + = Hedges' g

Effect Size Calculation

In this study, the results from multiple studies are converted into a quantitative measure called effect size (ES). The ES based Hedges' g is computed to determine the effect of feedback in CALL. Hedges' g is commonly used to adjust the difference in sample size to obtain unbiased estimates of effect size (Adesope et al., 2017; Borenstein et al., 2009; Kang & Han, 2015; Li, 2010). It shows the difference between the posttest means of experimental and control groups. This difference is shown by calculating the means (M) of both groups and dividing the means by the pooled standard deviation (SD) (Lipsey & Wilson, 2001). The investigator used CMA to calculate the ES . All studies except Nagata and Swisher (1995) provided the necessary statistics (M , SD , sample size) to calculate the ES . Nagata and Swisher (1995) did not provide M or SD . They reported t statistics. For outliers, "any absolute value (regardless of whether it was positive or negative) larger than 2.0 was eliminated from the analysis" (Li, 2010, p. 328). For the magnitude of the effect size based on the context of language research, "0.40 should be interpreted as small, 0.70 as medium, and 1.0 as large" (Kang & Han, 2015, p. 10).

Random-Effects Model

The random-effect model assumes that the effect sizes differ from one study to another. The choice to report the random-effects model for moderators is because the included studies had different experimental treatments and control conditions. They were carried out at different times, with different populations, and in different settings. There was no common population for these disparate studies. Therefore, the random-effect model fit the data collected for this review (Borenstein et al., 2009).

Evaluating Publication Bias

The visual inspection of the funnel plot was used to assess whether the publication bias exists visually. It is a scatterplot of the effect size from included studies against the measure of the study's precision. If the funnel plot is symmetric (i.e., equal distribution of the studies), this indicates there is no significant publication bias. On the other hand, if it is asymmetric, this means that there is possible publication bias (Belland et al., 2017; Borenstein et al., 2009).

The analysis included 21 studies. The majority of the studies' effect sizes ($k=11$) were not statistically significant (see Table 2). This could imply that there is a possible threat of validity in the results of this meta-analysis. Therefore, a funnel plot was created to examine whether publication bias exists.

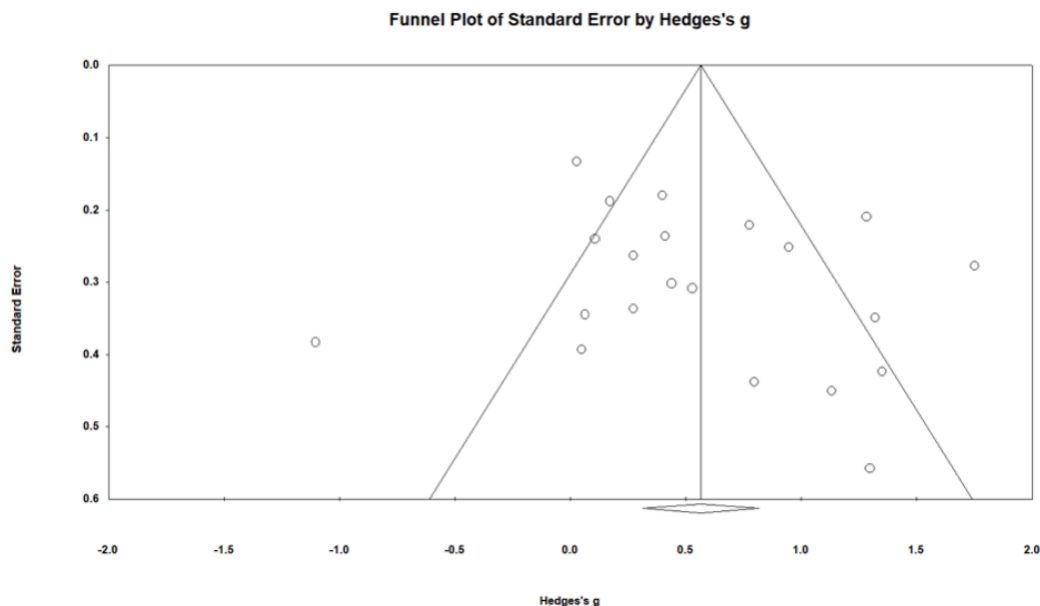


Figure 1. Immediate Posttest Funnel Plot.

Figure 1 shows that the plots are asymmetrical. Some effect sizes (i.e., plots) fall out of the funnel. In addition, there is an obvious absence of studies at the bottom on both sides. This could be evidence of the publication bias. To address this, a classic fail-safe N test was conducted. The result provided a statistically significant z -value ($z=9.02, p < .05$) indicating that (425) additional studies are needed to raise the p -value in order to have non-significant z -value. An Orwin's fail-safe N was performed with the 425 studies that must invalidate the overall effect size result and reject the null hypothesis that the effect size is the same as zero. The criterion trivial level was set at .05 and it showed that 183 studies were required to invalidate the overall effect. This number exceeded the criterion (i.e., $5k + 10$ limit) number (Adesope et al., 2017; Kang & Han, 2015). Therefore, the fail-safe N result indicated that publication bias was not an issue threatening the validity of the findings in this meta-analysis.

Results and Discussion

The results are presented in a sequence related to the two research questions.

Research Question 1: What is the overall effect of CALL feedback on language learning?

Table 3 shows that the overall weighted mean effect size was statistically significant ($RE: g = 0.56$), showing that feedback in CALL has a moderate and positive impact on students' language learning outcomes. With regard to the heterogeneity of the included studies, the analysis ($Q(20) = 93.43, p < .05, I^2 = 78.60$) showed a significant heterogeneity across the effect sizes of the studies implying that 78.60%, heterogeneity might be due to the differences across studies (Li, 2010; Kang & Han, 2015).

Table 3. The Overall Posttest Weighted Mean Effect Size of Feedback in CALL on Language Learning.

Model	K	Effect size			Test of heterogeneity			
		g+	SE	CI (95%)	Q	Df	P	I ²
Random	21	0.56	0.12	[0.31, 0.81]	93.43	20	0.00	78.60

Note. *= $p < .05$, K=number of studies, SE=standard error, CI= confidence interval, Df= degree of freedom, g+= Hedges' g

Previous meta-analyses varied in the magnitude of the effect size. In Li's (2010) analysis, the studies examined the effect of computer corrective feedback on the second language and generated a mean effect size of FE: $d = 0.61$ and RE: $d = 0.64$; Kang and Han (2015) analysed the effect of traditional written corrective feedback on grammatical accuracy and the study produced a moderate effect size (RE: $d = 0.68$). Van der Kleij et al. (2015) conducted an analysis investigating the effectiveness of feedback in computer instruction environments on student learning in different disciplines (e.g., math, psychology) rather than language learning. The analysis under the mixed model effects showed a significant effect of feedback ($d = 0.49$) on student learning. The differences among the aforementioned meta-analyses in the magnitude of effect size could be attributed to the variations in their (a) inclusion/exclusion criteria and (b) focus. This study differs from the previous meta-analyses in excluding any study in which the mother tongue and the target language are the same. This ensures that the goal is to learn a language and it is not a concept learning. Regardless of the criteria and the magnitude of the effect size, literature that investigated the effect of feedback on learning in general (Azevedo & Bernard, 1995; Van der Kleij et al., 2015) and language learning (Li, 2010; Kang & Han, 2015) is consistent with this study results that feedback in CALL could facilitate language learning.

Research Question 2: What are the variables that moderate the effect of CALL feedback on language learning?

Moderator analysis, using the random-effects model, was conducted to explore which variables were influencing the effect of feedback in CALL. The significance of any moderator variable was determined by the *Q*-statistics. The significant and insignificant moderator variables are presented and discussed separately below.

Significant Moderators

Table 4 shows the variation associated with participant educational level from three subcategories. The influence of participant educational level on feedback in CALL study was examined by categorized studies into three levels: Graduate level ($g = -1.34, k = 1, p < .05$), undergraduate ($g = 0.62, k = 17, p < .05$), and mixed level ($g = 0.74, k = 3, p < .05$). In addition, the influence of participant educational level ($Q(2) = 18.80, p < .05$) was found to significantly moderate the overall effect size of feedback in CALL. The significance of this moderator also means that the difference between these levels was significant and each level could have an impact on the effect of feedback on CALL.

Although graduate-level exhibited larger mean effect sizes compared to other levels, it is important to note that one out of 21 studies was with graduate students. Given this small number of studies with only graduate students, it is difficult to make a conclusive claim that the graduate level is more

influential on feedback in CALL compared with other levels. Regardless of the number of studies for each level, the difference between these levels was statistically significant.

Table 4. Random-Effects Model: Summary of Significant Moderators

Moderator	K	Effect size		CI (95%)		Test of heterogeneity	
		g+	SE	Lower	Upper	Q	Df
Educational Level							
Graduate	1	-1.10*	0.38	-1.85	-0.35		
Mixed	3	0.74	0.42	0.09	1.58		
Undergraduate	17	0.62*	0.13	0.40	0.87		
Between-Levels (Q _B)						18.80*	2
Intervention Provider							
Instructor	10	0.62*	0.21	0.19	1.05		
NR	1	0.03	0.13	-0.23	0.29		
Researcher	10	0.56*	0.16	0.24	0.88		
Between-Levels (Q _B)						9*	2
Mother Tongue							
Chinese	2	0.71	0.41	-0.11	1.53		
English	9	0.55*	0.11	0.33	0.76		
Japanese	1	0.02	0.13	-0.23	0.29		
Korean	1	1.13*	0.45	0.25	2.01		
Mixed	7	0.44	0.32	-0.18	0.08		
Persian	1	1.28*	0.20	0.87	0.69		
Between-Levels (Q _B)						28.99*	5
Research Context							
Classroom	11	0.57*	0.11	0.34	0.79		
Computer Lab	9	0.55*	0.27	0.01	1.09		
NR	1	0.02	0.13	-0.23	0.28		
Between-Levels (Q _B)						10.13*	2
Subject Domain							
Grammar	14	0.70*	0.14	0.42	0.97		
Listening Comprehension	1	0.41	0.23	-0.05	0.87		
Pronunciation	1	-1.10*	0.38	-1.85	0.35		
Reading Comprehension	1	0.03	0.13	-0.23	0.30		
Speaking	2	0.90*	0.40	0.11	1.68		
Writing	1	0.06	0.34	-0.61	0.73		
Between-Levels (Q _B)						28.40*	5
Target Language							
Dutch	1	-1.10*	0.38	-1.85	-0.35		
English	8	0.57*	0.20	0.17	0.97		
French	2	0.34	0.24	-0.13	0.82		
German	2	0.84	0.45	-0.04	1.73		
Japanese	3	0.29*	0.13	0.03	0.55		
Spanish	5	0.99*	0.24	0.51	1.47		
Between-Levels (Q _B)						23.77*	5

Note. * = $p < .05$, K=number of studies, SE=standard Error, CI= confidence interval, Df= degree of freedom, g+= Hedges' g

The data regarding the intervention provider are given in Table 4. The analysis showed that the intervention provider ($Q = 9, p < .05$) was a statistically significant moderator. An instructor as an intervention provider ($g = 0.62, k = 10, p < .05$) had a significantly higher influence on feedback

in CALL than a researcher ($g= 0.56, k= 10, p <.05$), and non-reported studies ($g= 0.03, p >.05$). A possible reason why intervention given by instructors has higher influence could be that while teachers use applications to address student needs, researchers test whether applications teach students an aspect of language (e.g., past tense). However, this explanation is a hypothesis of what the reason might be. Since there is no review available examined the influence of the intervention provider, research is needed to empirically test the intervention provider and the difference between its subcategories (e.g., instructor, researcher).

In the context of language influence, Table 4 indicated that the mother tongue ($I(5) = 28.99, p <.05$) significantly moderates the overall effect of feedback. In addition, the analysis shows that Persian ($g= 1.28, k= 1, p <.05$), had the highest effect on feedback, followed by English ($g= 0.55, k= 9, p <.05$), Korean ($g= 1.13, k= 1, p <.05$), Chinese ($g= 0.71, k= 2, p >.05$), mixed ($g= 0.44, k= 7, p >.05$), and Japanese ($g= 0.02, k= 1, p >.05$). Although the mother tongue moderates the effect of feedback, its effect has been understudied in previous meta-analyses. This new finding might be consistent with the mother tongue interference theory (Manan et al., 2017) that the mother tongue has an effect on target language. These results show that mother tongue could influence feedback in CALL as well.

Research context ($I(2) = 10.13, p <.05$) played a significant moderating role in the overall effect of feedback in CALL. The results suggested that research context had the potential to have a great effect when the intervention is given in classroom context ($g= 0.57, k= 11, p <.05$) and computer labs ($g= 0.55, k= 0, p <.05$) as well. This finding is not in line with Li (2010). One possible explanation is that Li (2010) had few studies where feedback was delivered by the computer.

Similarly, the effect sizes of the subgroups in the subject domain differed significantly, ($I(5) = 28.40, p <.05$). Table 4 indicated that pronunciation ($g= -1.10, k= 1, p <.05$) had higher influence on feedback than speaking ($g= 0.90, k= 2, p <.05$), grammar ($g= 0.70, k= 14, p <.05$), listening comprehension ($g= 0.41, k= 1, p >.05$), writing ($g= 0.06, k= 1, p >.05$), and reading comprehension ($g= 0.03, k= 1, p >.05$). As shown in Table 4, the subject domain was a statistically significant moderator of feedback in CALL. The size associated pronunciation had the language highest effect sizes. Since there is one study of pronunciation compared with many studies in speaking ($k= 2$), grammar ($k= 14$), and so on, it is difficult to make a conclusive claim about the upper hand of pronunciation.

For the target language, the effect sizes of 6 languages were computed to determine to what extent the effectiveness of feedback in CALL was dependent on the target language. Unlike Kang and Han (2015) and Li (2010), significant differences ($I(5) = 23.77, p <.05$) were found between Spanish ($g= 0.99, k= 5, p <.05$), English ($g= 0.57, k= 8, p <.05$), French ($g= 0.34, k= 2, p >.05$), German ($g= 0.84, k= 2, p >.05$), Japanese ($g= 0.29, k= 3, p <.05$), and Dutch ($g= -1.10, k= 1, p >.05$). The discrepancy between Kang and Han (2015) and Li (2010) in one side and this study on the other side about the potential influence of the target language might be due to differences in the inclusion criteria each study adopted. As shown in Table 4, the target language is a significant moderator of feedback in CALL.

Non-Significant Moderators

Table 5 did not find a significant amount of explained effect-size heterogeneity for the intervention length ($I(3) = 3.76, p >.05$). The long-term interventions ($g= 0.97, k= 3, p <.05$) had more effect than medium ($g= 0.54, k= 4, p >.05$), short ($g= 0.50, k= 13, p <.05$), and non-reported intervention length ($g= 0.94, k= 1, p >.05$). However, the Q -tests showed no significant difference between long-term, short-term, and medium interventions. This indicates that the length of the intervention does not have an impact on the overall effect of feedback in CALL.

Table 5. Random-effects Model: Summary of Insignificant Moderators.

Moderator	K	Effect size		CI (95%)		Test of heterogeneity	
		g+	SE	Lower	Upper	Q	Df
Intervention Length							
Long	3	0.97*	0.34	0.29	1.65		
Medium	4	0.54	0.28	-0.01	1.11		
NR	1	0.94*	0.25	0.45	1.43		
Short	13	0.50*	0.17	0.11	0.78		
Between-Levels (QB)						3.76	3
Intervention Modeling							
Given	12	0.59*	0.21	0.18	1		
NR	9	0.48*	0.13	0.22	0.75		
Between-Levels (QB)						0.17	1
Intervention Status							
Developed	13	0.58*	0.17	0.25	0.92		
Literature	4	0.68*	0.29	0.10	1.26		
NR	4	0.39	0.30	0.21	0.99		
Between-Levels (QB)						0.47	2
Language Proficiency							
Advanced	1	0.77*	0.22	0.34	1.21		
Beginner	5	0.07	0.26	-0.44	0.59		
Intermediate	9	0.73*	0.19	0.35	1.11		
Mixed	6	0.65*	0.25	0.15	1.16		
Between-Levels (QB)						5.08	3
Measures of Proficiency							
Class Enrolment	1	0.06	0.34	-0.61	-0.73		
Placement Test	2	0.54	0.62	-0.67	1.77		
Pretest	18	0.60*	0.13	0.33	0.87		
Between-Levels (QB)						2.11	2
Publication Type							
Article	12	0.50*	0.17	0.15	0.84		
Dissertation	8	0.67*	0.22	0.24	1.11		
Master's Thesis	1	0.44	0.30	-0.15	1.03		
Between-Levels (QB)						0.53	2
Research Setting							
Foreign	18	0.59*	0.14	0.31	0.87		
Second	3	0.35	0.26	-0.16	0.87		
Between-Levels (QB)						0.62	1

Note. *= $p < .05$, K=number of studies, SE=standard Error, CI= confidence interval, Df= degree of freedom, g+= Hedges' g

Another moderator is the intervention modeling. It is one of the few moderators that have never been explored. Table 5 did not show a significant effect-size heterogeneity for intervention modeling ((1) = 0.17 $p > .05$). This indicates that effect sizes between the modeled intervention(s) (i.e., participant receiving training on how to use the software application(s) and the non-reported

(i.e., no mentioning to whether the application is modeled) did not significantly differ. Although no statistical differences were found, it is noteworthy to highlight that modeled intervention(s) ($g=0.58, k=12, p < .05$) could have a considerable influence on the effect of feedback in CALL than the non-modeled intervention ($g=0.48, k=8, p < .05$).

With regard to the intervention status ($I(2) = 55.52$), Table 5 showed that it did not appear to be an influential moderator. This demonstrates that there were no significant differences in subcategories: The literature-based interventions ($g=0.68, k=3, p < .05$), developed interventions ($g=0.58, k=12, p < .05$) and non-reported studies ($g=0.39, k=3, p > .05$). Nevertheless, literature-based interventions have the highest impact on the feedback effect. One explanation could be that these interventions have been structured and validated based on feedback literature. However, this moderator and its new findings have never been investigated in previous studies. Therefore, future research should examine the status of the intervention in detail.

For the language proficiency, the data regarding the language proficiency ($I(3) = 5.08, p > .05$) showed no statistical difference between advanced language proficiency ($g=0.77, k=1, p < .05$) intermediate ($g=0.74, k=9, p < .05$), beginner ($g=0.07, k=5, p > .05$), and mixed ($g=0.65, k=6, p < .05$). Different from Kang and Han (2015), who found that language proficiency is a significant moderator; this meta-analysis revealed language proficiency was not as a moderator variable. Although it was not statistically significant, this result of this study is in line with the claim by Kang and Han (2015) that advanced language proficiency could have more influence on feedback than the intermediate, beginner, and mixed-level language proficiency. As shown in Table 5, there are no statistical differences between proficiency levels. More research is needed to investigate the reason for the ineffectiveness of the proficiency level on feedback effect in CALL.

With respect to the measures of proficiency ($I(2) = 2.11, p > .05$), it was not a significant moderator of the CALL feedback effect. However, pretest ($g=0.60, k=18, p < .05$) was significantly higher than class enrolment ($g=0.06, k=1, p > .05$) and placement test ($g=-0.54, k=2, p > .05$). The analysis revealed there were no significant differences between measures of proficiency. Therefore, the measure of proficiency did not play any moderating role in the effect of feedback.

Similar to previous meta-analysis (Li, 2010) in the field, this study showed that publication type (see Table 5) was not a statistically significant moderator ($I(2) = 0.53, p > .05$). Although dissertations ($g=0.67, k=8, p < .05$) yielded a significantly larger effect size than articles ($g=0.50, k=12, p < .05$) and Master's thesis ($g=0.44, k=1, p > .05$), the difference was not significant. According to Li 2010, this indicates there is no evidence of publication bias (i.e., the tendency to include studies in the analysis with only significant results).

As noted in Table 5, there were no significant differences between foreign language setting ($g=0.59, k=18, p < .05$) and second research settings ($g=0.35, k=3, p > .05$). Similar to Li (2010), foreign language settings produced a larger effect size than second language settings. It should be noted this aligns with Li's (2010) study and differs from Kang and Han's (2015). One explanation could be that Li (2010) includes few studies where feedback was given by the computer. However, Kang and Han (2015) excluded any study where feedback was delivered by the computer. To sum up, the research setting ($I(1) = 0.62, p > .05$) does not affect feedback in CALL on student language learning.

Conclusion

The present study quantitatively synthesized the results of 21 studies with a total of 1313 students about the overall effect of feedback in CALL on student learning outcomes. The results of this study are in line with previous meta-analyses (e.g., Azevedo & Bernard, 1995; Kang & Han 2015;

Li, 2010) in the field that feedback in CALL has a significant moderate and positive impact on language learning. In addition, the study investigated the factors moderating the effect of feedback in CALL. Among the thirteen moderator variables analyzed, educational level, intervention provider, mother tongue, research context, subject domain, and target language had a significant impact on the overall effect of feedback. However, factors including intervention length, intervention modeling, intervention status, language proficiency, measures of proficiency, publication type, and research setting did not show an effect on feedback in CALL.

Implication

This study unlocks salient findings for teachers interested in feedback in CALL. By better understanding how the moderators (i.e., related to application characteristics such as modeling) work before starting to use applications in teaching, teachers can take into consideration these factors to make a choice using this study results as criteria to select applications they need for their teaching. In addition, instructors could predict students' performance results based on learners' characteristics (e.g., educational level, mother tongue). Therefore, instructors can better employ feedback in CALL to fulfill their students' needs.

For researchers, this study provides important results and new findings. It is expected to help scholars identify new areas for research. For example, one area, where research is needed, is the mother tongue. The results showed participant mother tongue had a significant effect. However, the level of language proficiency did not show any effect. It seems that although students might have a good level of language proficiency, their mother tongue could still affect their understanding of feedback in CALL. Future research should examine the effect of the mother tongue on feedback in CALL.

In addition, the study synthesized quantitative research on feedback in CALL. However, this study did not include qualitative or quantitative studies without a control or comparison group(s). Therefore, the results of this study do not reflect all empirical research of feedback in CALL. A systematic review is needed to summarize qualitative and non-control comparison quantitative studies. These studies are beneficial to provide insight into the effect of feedback in CALL. Moreover, the study identified the dearth of studies with young learners at primary and secondary schools. Future research should explore the effect of feedback in CALL with participants at primary and secondary schools.

About the Author

Adnan Mohamed is a Ph.D. candidate in Language, Literacy, and Technology at Washington State University. His research interests include computer-assisted language learning and feedback

References

References marked with an asterisk indicate studies included in the analysis.

- Adesope, O. O., Trevisan, D., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*, 659-701.
- Ai, H. (2017). Providing graduated corrective feedback in an intelligent computer-assisted language learning environment. *ReCALL, 29*(3), 313-334.
- Anh, D. T. K., & Marginson, S. (2013). Global learning through the lens of Vygotskian sociocultural theory. *Critical Studies in Education, 54*(2), 143-159.
- Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research, 13*, 111-127.
- *Bationo, B. (1992). The effects of three forms of immediate feedback on learning intellectual skills in a foreign language computer-based tutorial. Retrieved from <http://search.proquest.com/docview/85554616/>
- Beechler, S., & Williams, S. (2012). Computer assisted instruction and elementary ESL students in sight word recognition. *International Journal of Business and Social Science, 3*(4), 85-92.
- Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in stem education: A meta-analysis. *Review of Educational Research, 87*, 309-344.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis* (Version 2.2.027) [Computer software]. Biostat.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein H. R. (2009). *Introduction to meta-analysis*. Wiley.
- Cornell, J., & Mulrow, C. (1999). Meta-Analysis. In A. Herman & M. Gideon (Eds.), *Research mythology in the social, behavioral, and life sciences* (pp. 285-323). SAGE Publications.
- Dongyu, Z., Fanyu, & Wanyi, D. (2013). Sociocultural theory applied to second language learning: Collaborative learning with reference to the Chinese context. *International Education Studies, 6*(9), 165-174.
- Duhon, G., House, S., Hastings, K., Poncy, B., & Solomon, B. (2015). Adding immediate feedback to explicit timing: an option for enhancing treatment intensity to improve mathematics fluency. *Journal of Behavioral Education, 24*(1), 74-87.
- Elola, I., & Oskoz, A. (2016). Supporting second language writing using multimodal feedback. *Foreign Language Annals, 49*(1), 58-74.
- Englert, C.S., Mariage, T.V., & Dunsmore, K. (2006). Tenets of sociocultural theory in writing instruction research. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 208- 221). Guilford.
- Fahim, M., & Haghani, M. (2012). Sociocultural perspectives on foreign language learning. *Journal of Language Teaching & Research, 3*(4), 693-699.
- Fajfar, P., Campitelli, G., & Labollita, M. (2012). Effects of immediacy of feedback on estimations and performance. *Australian Journal of Psychology, 64*, 169-177.

- *Gomez, M. (2016). The effects of type of feedback, amount of feedback and task essentialness in a L2 computer-assisted study. ProQuest Dissertations Publishing. Retrieved from <http://search.proquest.com/docview/1853103093/>
- *Hanson, R. (2008). Feedback in intelligent computer-assisted language learning and second language acquisition: A study of its effect on the acquisition of French past tense aspect using an intelligent language tutoring system. ProQuest Dissertations Publishing. Retrieved from <http://search.proquest.com/docview/1685019938/>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Heift, T. (2010). Prompting in CALL: A longitudinal study of learner uptake. *Modern Language Journal*, 94(2), 198-216.
- *Heift, T., & Rimrott, A. (2008). Learner responses to corrective feedback for spelling errors in CALL. *System*, 36(2), 196-213.
- *Hincks, R., & Edlund, J. (2009). Promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning & Technology*, 13(3), 32-50.
- *Hsieh, H.-C. (2007). Input-based practice, feedback, awareness and L2 development through a computerized task. Retrieved from <http://search.proquest.com/docview/85686774/>
- Hung, S. (2016). Enhancing feedback provision through multimodal video technology. *Computers & Education*, 98, 90-101.
- Ifenthaler, D. (2010). Bridging the gap between expert-novice differences. *Journal of Research on Technology in Education*, 43(2), 103–117.
- Ilies, R., Judge, T., & Wagner, D. (2010). The influence of cognitive and affective reactions to feedback on subsequent goals: Role of behavioral inhibition/activation. *European Psychologist*, 15(2), 121-131.
- Kang, E, & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *Modern Language Journal*, 99(1), 1-18.
- Kargar, N., & Tayebipour, F. (2015). The effect of scaffolding on EFL learners' reading comprehension. *Modern Journal of Language Teaching Methods*, 5(4), 446-453.
- *Kim, D.-H. (2009). Explicitness in CALL feedback for enhancing advanced ESL learners' grammar skills. ProQuest Dissertations Publishing. Retrieved from <http://search.proquest.com/docview/85706442/>
- *Kregar, S. (2011). Relative effectiveness of corrective feedback types in computer-assisted language learning. Retrieved from <http://search.proquest.com/docview/1081897605/>
- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60(2), 309-365.
- *Lavolette, E. (2014). Effects of feedback timing and type on learning ESL grammar rules. ProQuest Dissertations Publishing. Retrieved from <http://search.proquest.com/docview/1553437096/>
- *Lavolette, Elizabeth, Polio, Charlene, & Kahng, Jimin. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language Learning & Technology*, 19(2), 50-68.

- *Lee, S.-P., Su, H.-K., & Lee, S.-D. (2012). Effects of computer-based immediate feedback on foreign language listening comprehension and test-associated anxiety. *Perceptual & Motor Skills*, 114(3), 995–1006.
- Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: SAGE Publications.
- Liutkus, D. (2012). Investigating CALL in the classroom: situational variables to consider. *Advances in Language and Literary Studies*, 3(1), 107-118.
- Manan, N. A. A., Zamari, Z. M., Pillay, I. A., Adnan, A. H. M., Yusof, J., & Raslee, N. N. (2017). Mother tongue interference in the writing of English as a second language (ESL) Malay learners. *International Journal of Academic Research in Business and Social Sciences*, 7(11), 1294-1301.
- Masuda, K., Arnett, Carlee, & Labarca, Angela. (2015). *Cognitive linguistics and sociocultural theory: Applications for second and foreign language teaching* (Studies in Second and Foreign Language Education [SSFLE]). De Gruyter Mouton.
- Meo, S. (2013). Giving feedback in medical teaching: A case of lung function laboratory/spirometry. *College of Physicians & Surgeons Pakistan*, 23(1), 86–89.
- *Moreno, N. (2007). The effects of type of task and type of feedback on L2 development in CALL. Published doctoral dissertation, Georgetown University, Washington, DC.
- *Murphy, P. (2007). Reading comprehension exercises online: The effects of feedback, proficiency and interaction. *Language Learning & Technology*, 11(3), 107-129.
- *Nagata, N. (1993). Intelligent computer feedback for second language instruction. *Modern Language Journal*, 77(3), 330–339.
- *Nagata, N., & Swisher, M. V. (1995). A Study of consciousness-raising by computer: The effect of metalinguistic feedback on second language learning. *Foreign Language Annals*, 28(3), 337–347.
- *Neri, A., Cucchiarini, C., & Strik, H. (2008). The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. *ReCALL*, 20(2), 225–243.
- *Park, D., (1998). *Focus on form: The role of negative feedback in German relative pronoun selection*. (Unpublished master's thesis). Michigan State University, East Lansing, Michigan.
- *Rashidi, N., & Babaie, H. (2013). Elicitation, recast, and meta-linguistic feedback in form-focused exchanges: Effects of feedback modality on Multimedia Grammar Instruction. *The Journal of Teaching Language Skills*, 4 (4), 25-51.
- *Sachs, R. (2012). Individual differences and the effectiveness of visual feedback on reflexive binding in L2 Japanese. Retrieved from <http://search.proquest.com/docview/1018383321/>
- *Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning*, 54(1), 35–78.
- *Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language, Learning & Technology*, 13(1), 96-120.
- Stickler, U., & Shi, L. (2016). TELL us about CALL: An introduction to the virtual special issue (VSI) on the development of technology enhanced and computer assisted language learning published in the System Journal. *System*, 56, 119-126.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, (2), 50-73.

Van der Kleij, F., Feskens, R., & Eggen, T. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: *A Meta-Analysis. Review of Educational Research*, 85(4), 475-511.

Van Laere, E., Agirdag, O., & Van Braak, J. (2016). Supporting science learning in linguistically diverse classrooms: Factors related to the use of bilingual content in a computer-based learning environment. *Computers in Human Behavior*, 57(C), 428-441.

Wang, M., Wu, B., Kirschner, P., & Spector, M. (2018). Using cognitive mapping to foster deeper learning with complex problems in a computer-based environment. *Computers in Human Behavior*, 87, 450-458.

Copyright rests with authors. Please cite TESL-EJ appropriately.

Copyright © 1994 - 2020 TESL-EJ, ISSN 1072-4303