The Effectiveness of Explicit Corrective Feedback in the Second Language Classroom

Paul Chamness Miller, Akita International University
Wei Pan, Duke University

Abstract

Despite an increasing push for standardisation in schools, where the focus is increasingly turning to testing, language classrooms in some locations around the world (e.g., North America and Japan) are moving away from grammar-emphasized teaching filled with drills, for an approach that is more communicative (see Lee & VanPatten, 2003; Omaggio Hadley, 2000; Richards, 2006). Research suggests that exposure to native-like language is not enough for learners to achieve acquisition (White, 1991). Student, instead, require opportunities to produce language, complete with errors. For language teachers, the question is how to respond to students' errors in the classroom, in particular to consider whether error correction is effective. Although many studies have examined this problem, the results are mixed. In order to sort out these mixed results, this study presents the results of a meta-analysis on the effects of explicit oral corrective feedback in the second language classroom. The results reveal a small effect on the ability of explicit feedback to promote language learning, although whether there is a long-term effect that leads to acquisition is not yet confirmed.

Keywords: second language learning; corrective feedback; explicit feedback; second language acquisition; meta-analysis; foreign language teaching

Introduction

Despite an increasing push for standardisation in schools, where the focus is increasingly turning to testing, some second language (L2) classrooms around the globe are moving away from grammar-emphasised teaching filled with drills, for an approach that is more communicative. For example, North American countries have been moving in this direction for decades, with communication being at the centre of guidelines issued by the American Council on the Teaching of Foreign Languages (ACTFL) (1996), and other countries are also revising their language curriculum to centre on communication. One such example is Japan, whose Ministry of Education, Culture, Sports, Science and Technology (MEXT) (2008), has recently

revised the foreign language objective so that it reads: “To form the foundation of pupils’ communication abilities through foreign languages while developing the understanding of languages and cultures through various experiences, fostering a positive attitude toward communication, and familiarizing pupils with the sounds and basic expressions of foreign languages.”

In order to achieve the goals and objectives of a communicative approach, second language teaching requires student interaction with peers and teachers in the second language. This approach to language teaching challenges the traditional practices one commonly found in the classroom in the past, such as the audiolingual method, an approach that is filled with drills and worksheets (see Richards, 2006, for a more detailed description of the evolution of language teaching). Through the work of Chomsky (1959), we now understand that behaviourist teaching methodology does not lead to true language learning or acquisition. Instead, Hymes (1966, 1972) pioneered the way toward understanding that language learning is much more than memorizing rules, but instead is a complex system of competencies.

One of the primary questions that teachers still ask today is how to develop these competencies in language learners while promoting language accuracy. We have learned that positive evidence (i.e., exposure to correct language use) is not enough for learners to achieve acquisition (White, 1991). In order to acquire language, students require opportunities to produce language, complete with errors, and to have occasions for self-correction through interaction with others (Swain, 1985, 1989). It is in this interaction that learners will begin to notice their errors and modify their language production. However, language teachers must make decisions as to how to raise learners’ consciousness. One common method handed down from one generation of teachers to another is responding to students’ errors in the classroom in a corrective manner (i.e., corrective feedback). However, the question remains, is error correction effective and necessary, and if so, which methods of addressing these errors are most effective? Unfortunately, this question does not have a simple answer, as many factors appear to play a role in addressing errors (see DeKeyser, 1993; Long, 2007). Some studies suggest that error correction has no effect on language acquisition (see Carroll, 1999; Truscott, 1999), while others, such as Schmidt (1990), have argued that corrective feedback result in learners *noticing* errors, leading to potential acquisition. Despite the mixed results of research, the common practice among language teachers is the need to correct students’ errors in one

form or another. Furthermore, many studies have found that students expect their teachers to correct errors (see, for example, Diab, 2005; Ji, 2015; Papangkorn, 2015; Salteh & Sadeghi, 2015).

There are a number of approaches to correcting errors, such as *recasts* (where the teacher implicitly restates an incorrect utterance correctly) and *explicit feedback*, which can either be simply telling the learner that what he or she said was wrong, or offering more a more metalinguistic explanation as to what is wrong and why or helping the student recall a rule. The many options that language teachers have in treating errors raises questions in determining how to help their students best achieve acquisition. However, the literature reveals mixed results, failing to provide language teachers definitive results from which to make informed decisions in their classrooms. For example, both DeKeyser (1993) and Long (2007) have noted that there are too many variables that are connected to the potential effectiveness of error correction (e.g., age, linguistic proficiency, cognitive ability, personality, among others). Others have even argued that error correction has no impact on error correction (see Carroll, 1999; Truscott, 1999). Other studies suggest that although error correction may not directly lead to acquisition, it may raise their consciousness to notice errors (Schmidt, 1990).

Despite this lack of a clear indication about which type of feedback is most effective, many language teachers default to explicit error correction, as evidenced by the many studies that have examined explicit error correction (for example, Leow, 1998; White, Spada, Lightbown, & Ranta, 1991; Yilmaz, 2012). In order to understand what we know about the effectiveness of explicit feedback, and to sort out the conflicting results of the studies, we have conducted a meta-analytic review of the current research on explicit error correction in the L2 classroom in an effort to answer the question, does explicit corrective feedback have a positive effect on the learning of a second language?

Method

Literature Search

A preliminary literature search using ERIC, JSTOR, the Social Sciences Citation Index, and Dissertation Abstracts International produced approximately 450 relevant articles and dissertations published between 1980 and 2013, with keywords of “explicit,” “corrective,” “feedback,” and “negative evidence,” and all the possible combinations of these key words. An

ancestry search in the initial articles resulted in approximately 30 additional articles not already in the original search that we considered for possible inclusion in the study. After applying the criteria for inclusion described in the following section, there were 71 studies in the 22 final articles that we were able to include in this meta-analysis.

Criteria for Inclusion

The first criterion for inclusion in the present meta-analysis was that the studies measure the effect of explicit corrective feedback on some aspect of a second or foreign language. Additionally, we only included randomized or quasi-experiments, and each study was required to report all of the appropriate information, including the sample size, mean score and standard deviation of both the control and experimental groups. In cases where portions of this information were missing, we contacted the author(s) in an effort to obtain the missing information. In one instance, we were unable to obtain the missing data and consequently excluded the study from the analysis.

Recorded Study Characteristics

We included in our analysis the first and second language of the participants, the linguistic structure that was the target of the study (e.g., knowing when to use simple past tense (preterit) or imperfect), the type of treatment implemented in the study (e.g., information gap), type of experimental group (e.g., randomized, in-tact, etc.), the type of treatment activities (e.g., oral role-play), length of study from pre- to final post-test, and the number of treatment sessions. Where there was more than one post-test, we only included the last post-test in our analysis. There were additional characteristics that we initially noted, but later deemed inappropriate for analysis at this time due to missing information or design of the studies in question. In particular, we excluded age and gender as the majority of the studies did not report this data.

Outcome Measure

For the present meta-analytic review, the outcome measure was whether explicit corrective feedback has an effect on SLA. The effect size Δ (Becker, 1988; Carlson & Schmidt, 1999; Morris & DeShon, 2002) of explicit corrective feedback on SLA was computed as the difference between the standardized mean change from the pretest to the posttest (or the delayed posttest where available) for the treatment and control groups, relative to the pooled standard deviation at the posttest or the delayed posttest:

$$\Delta = \frac{(\mu_{T/post} - \mu_{T/pre}) - (\mu_{C/post} - \mu_{C/pre})}{\sigma_{pooled/post}}$$

Coding

In order to address the reliability of the coding, we performed concurrent double coding by asking a colleague who also has a background in second language acquisition to code independently of the researchers approximately one third of the studies. The coding consisted of identifying each of the variables listed in Table 1 using a blank spreadsheet similar to this table. Additional coding involved using a checklist containing the inclusion criteria to evaluate the eligibility of these same studies. When completed, we compared our own coding, table by table and checklist by checklist with that of the independent coder. The result of this procedure was 100% agreement between raters in all characteristics of each study evaluated. Because there was complete agreement, no inter-rater reliability score was calculated.

Table 1. *Characteristics of the Studies*

Study	Setting	L1	L2	Structure	Design	Mode of Treatment
Master(1994)A	College	Mixture	English	Articles	Intact	Oral & Written
Master(1994)B	College	Mixture	English	Articles	Intact	Oral & Written
Alanen(1995)A	Lab	Mixture	Finnish	Locative Suffixes	Random	Oral & Written
Alanen(1995)B	Lab	Mixture	Finnish	Consonant Changes	Random	Oral & Written
Yilmaz (2012) A	Lab	English	Turkish	Plural morpheme	Random	Oral & Written
Yilmaz (2012) B	Lab	English	Turkish	Plural morpheme	Random	Oral & Written
Yilmaz (2012) C	Lab	English	Turkish	Plural morpheme	Random	Oral & Written
SBS (2012) A	Lab	Spanish	Latin	Morphosyntax	Random	Computer
SBS (2012) B	Lab	Spanish	Latin	Morphosyntax	Random	Computer
SBS (2012) C	Lab	Spanish	Latin	Morphosyntax	Random	Computer
SBS (2012) D	Lab	Spanish	Latin	Morphosyntax	Random	Computer
WE A	College	Mixture	English	Participal Adjectives	Intact	Oral & Written
WE B	College	Mixture	English	Participal Adjectives	Intact	Oral & Written
WE C	College	Mixture	English	Passive	Intact	Oral & Written
WE D	College	Mixture	English	Passive	Intact	Oral & Written
Nagata (1995)	Lab	Mixture	Japanese	Particles	Random	Computer
Nagata (1997b) A	Lab	Mixture	Japanese	Particles	Random	Computer
Nagata (1997b) B	Lab	Mixture	Japanese	Particles	Random	Computer
Kubota (1994) A	College	Japanese	English	Dative Alternation	Intact	Unknown

Study	Setting	L1	L2	Structure	Design	Mode of Treatment
Kubota (1994) B	College	Japanese	English	Dative Alternation	Intact	Unknown
Kubota (1994) C	College	Japanese	English	Dative Alternation	Intact	Unknown
Kubota (1994) D	College	Japanese	English	Dative Alternation	Intact	Unknown
Kubota (1994) E	College	Japanese	English	Dative Alternation	Intact	Unknown
Kubota (1994) F	College	Japanese	English	Dative Alternation	Intact	Unknown
de Graaf (1997) A	Lab	Dutch	eXperanto	Simple Morphology	Random	Computer
de Graaf (1997) B	Lab	Dutch	eXperanto	Complex Morpholoy	Random	Computer
de Graaf (1997) C	Lab	Dutch	eXperanto	Simple Syntax	Random	Computer
de Graaf (1997) D	Lab	Dutch	eXperanto	Complex Syntax	Random	Computer
de Graaf (1997) E	Lab	Dutch	eXperanto	Simple Morphology	Random	Computer
de Graaf (1997) F	Lab	Dutch	eXperanto	Complex Morpholoy	Random	Computer
de Graaf (1997) G	Lab	Dutch	eXperanto	Simple Syntax	Random	Computer
de Graaf (1997) H	Lab	Dutch	eXperanto	Complex Syntax	Random	Computer
de Graaf (1997) I	Lab	Dutch	eXperanto	Simple Morphology	Random	Computer
de Graaf (1997) J	Lab	Dutch	eXperanto	Complex Morpholoy	Random	Computer
de Graaf (1997) K	Lab	Dutch	eXperanto	Simple Syntax	Random	Computer
de Graaf (1997) L	Lab	Dutch	eXperanto	Complex Syntax	Random	Computer
de Graaf (1997) M	Lab	Dutch	eXperanto	Simple Morphology	Random	Computer
de Graaf (1997) N	Lab	Dutch	eXperanto	Complex Morpholoy	Random	Computer
de Graaf (1997) O	Lab	Dutch	eXperanto	Simple Syntax	Random	Computer
de Graaf (1997) P	Lab	Dutch	eXperanto	Complex Syntax	Random	Computer
Salaberry (1997)	College	English	Spanish	Direct Object	Intact	Oral & Written
Leow (1998) A	College	Mixture	Spanish	Preterit	Random	Oral & Written
Leow (1998) B	College	Mixture	Spanish	Preterit	Random	Oral & Written
Leow (1998) C	College	Mixture	Spanish	Preterit	Random	Oral & Written
Leow (1998) D	College	Mixture	Spanish	Preterit	Random	Oral & Written
CS (1993) A	College	Spanish	English	Dative Alternation	Random	Oral & Written
CS (1993) B	Lab	Spanish	English	Dative Alternation	Random	Oral & Written
DeKeyser (1995)	Lab	Mixture	Implexan	Grammatical Rules	Random	Computer
Doughty (1991)	Lab	Mixture	English	Relative clauses	Random	Computer
VP & C(1993) A	College	Mixture	Spanish	Direct Object	Intact	Oral & Written
VP & C (1993) B	College	Mixture	Spanish	Direct Object	Intact	Oral & Written
VP & O(1996) C	High School	English	Spanish	Direct Object	Intact	Oral & Written
VP & O(1996) D	High School	English	Spanish	Direct Object	Intact	Oral & Written
Kang (2009) A	Lab	English	Korean	Past Tense	Random	Oral & Written

Study	Setting	L1	L2	Structure	Design	Mode of Treatment
Kang (2009) B	Lab	English	Korean	Past Tense	Random	Oral & Written
Lyddon (2011)	Lab	Mixture	French	Preposition	Intact	Computer
Nagata (1997a)	Lab	Mixture	Japanese	Particles	Random	Computer
WSLR (1991) A	Primary	French	English	Question Formation	Intact	Oral & Written
WSLR (1991) B	Primary	French	English	Question Formation	Intact	Oral & Written
Lyster (1994) A	Junior High	English	French	Sociolinguistics	Intact	Oral & Written
Lyster (1994) B	Junior High	English	French	Sociolinguistics	Intact	Oral & Written
Lyster (1994) C	Junior High	English	French	Sociolinguistics	Intact	Oral & Written
Lyster (1994) D	Junior High	English	French	Sociolinguistics	Intact	Oral & Written
Lyster (1994) E	Junior High	English	French	Sociolinguistics	Intact	Oral & Written
Lyster (1994) F	Junior High	English	French	Sociolinguistics	Intact	Oral & Written
Lyster (1994) G	Junior High	English	French	Sociolinguistics	Intact	Oral & Written
Lyster (1994) H	Junior High	English	French	Sociolinguistics	Intact	Oral & Written
D&S (1996) A	College	Mixture	Spanish	Direct Object	Intact	Oral & Written
D&S (1996) B	College	Mixture	Spanish	Direct Object	Intact	Oral & Written
D&S (1996) C	College	Mixture	Spanish	Conditional	Intact	Oral & Written
D&S (1996) D	College	Mixture	Spanish	Conditional	Intact	Oral & Written

Statistical Analysis

Descriptive analysis was first conducted by calculating the unweighted mean effect size of the effect of explicit oral feedback on SLA, along with 95% confidence intervals of the effect sizes. The confidence intervals show whether the effect sizes are heterogeneous across the studies. In addition, a funnel plot of the effect sizes against the sample sizes was created to detect potential publication bias (Light & Pillemer, 1984). Typically, small studies tend to show more variability among the effect sizes than larger studies, and therefore, the plot should look like a funnel. A bite will be taken off from the funnel plot when publication bias exists against studies with small effect sizes (Greenhouse & Iyenger, 1994).

For inferential analysis, a Hedges and Olkin (1985) *Q*-statistic was computed. A test for the *Q*-statistic can provide statistical evidence for the heterogeneity of the 71 studies included in the meta-analysis. If the test is significant, a random-effects model is tested, and the weighted mean effect size is calculated to compare it with the unweighted mean effect size. In the case of heterogeneous effect sizes, the study characteristics are modeled in *ANOVA-like modeling*

(Hedges, 1994). ANOVA-like models are analogues to the regular ANOVA and they estimate the explained variation in the heterogeneous effect sizes based on the study characteristics. Following Hedges' (1994) suggestion, the standard error is adjusted as follows:

$$S_j = \frac{SE_j}{\sqrt{MS_{\text{Error}}}}$$

where *Sj* is the adjusted standard error, *SEj* is the original standard error as given by common computer programs, and *MS*Error is the mean square value for errors from the analysis of variance as given by the computer programs. It is also worth noting here that there were no missing data in the present meta-analytic study.

Results

Descriptive Analysis

The initial unweighted mean effect size was 0.27, ranging from -1.92 to 3.19. An initial confidence interval plot (Figure 1) shows that there were heterogeneous data on the effect sizes across the studies. Figure 2 is a funnel plot of the effect sizes against the sample sizes, which does not suggest possible publication bias, since no bite was taken off from the funnel.

Figure 1. 95% Confidence intervals of effect sizes

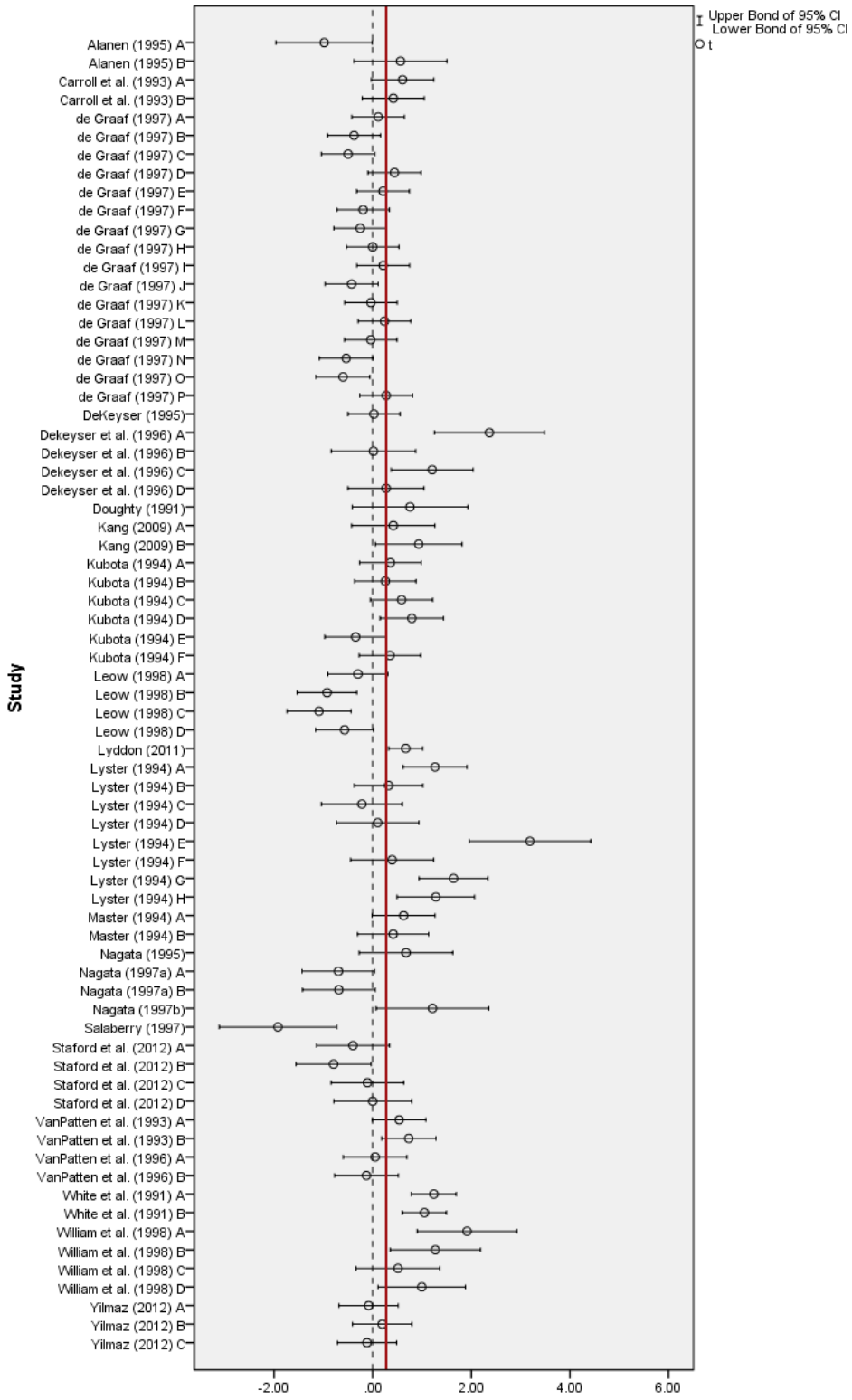
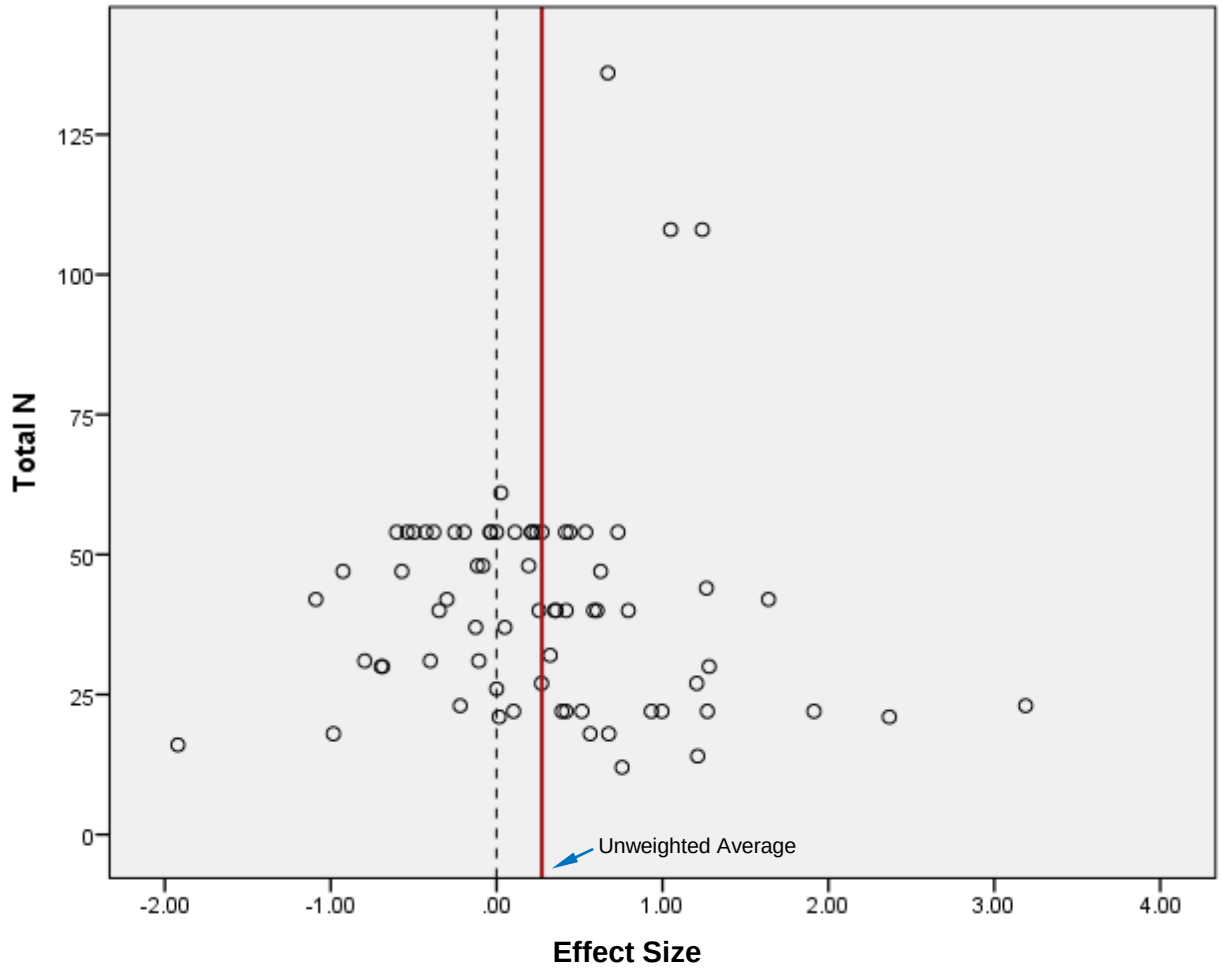


Figure 2. Funnel plot of effect sizes against sample sizes.



Inferential Analysis

Test of homogeneity. Under the null hypothesis of *H*0: θ1 = ... = θ71 = θ, where θ*i* (*i* = 1, ..., 71) and θ are the population parameters, the Hedges and Olkin (1985) *Q*-statistic value of *Q*Total was 284.13 with *df* = 70 ($p < .001$), which means that the effect sizes across the 71 studies were statistically heterogeneous. By testing a random-effects model under *H*0: θ = 0, a *z*-statistic = 2.58 ($p < .01$) indicates that the random-effects average of the 71 studies was statistically different from zero, and the weighted random-effects average was 0.25. Note that the unweighted average of the effect sizes was 0.27. Therefore, the 71 studies confirmed the actual success of the intervention of explicit oral correction on SLA.

ANOVA-like modelling. The ANOVA-like modelling was conducted for all recorded variables and found that setting, L1, L2, structure, design, and mode of treatment are significant factors

explaining the population variation of the effect size.

***Setting*.** As we can see in Table 2, the effect sizes between groups (i.e., setting) were heterogeneous. The between-group variation was significant ($Q_{between}$ = 72.84, *df* = 4, *p* < .001). Both the total within-group variation (Q_{within} = 211.29, *df* = 66, *p* < .001) and most of the within-group variations were significant (see Table 2). The results suggest that setting explained some (72.84/284.13 = 25.6%) of the population variation of the effect size.

Table 2. *Test Results of ANOVA-Like Model on Setting*

Source	*Q*-Statistic	*df*	*p*	Mean *ES*	*SE*
Between groups	72.84	4	< .001		
Within groups	211.29	66	< .001		
College	106.02	23	< .001	.27	.07
High School	.14	1	.706	-.04	.23
Junior High	34.31	7	< .001	.88	.14
Lab	70.48	34	< .001	.01	.05
Primary	.34	1	.559	1.14	.16
Total	284.13	70	< .001		

First language (L1). Table 3 shows that the effect sizes between groups (i.e., L1) were heterogeneous. The between-group variation was significant ($Q_{between}$ = 58.77, *df* = 5, *p* < .001). Both the total within-group variation (Q_{within} = 225.37, *df* = 65, *p* < .001) and some of the within-group variations were significant (Table 2). The results suggest that L1 explained some (58.77/284.13 = 20.7%) of the population variation of the effect size.

Table 3. *Test Results of ANOVA-Like Model on L1*

Source	*Q-Statistic*	*df*	*p*	Mean *ES*	*SE*
Between groups	58.77	5	< .001		
Within groups	225.37	65	< .001		
Dutch	21.43	15	.124	-.09	.07
English	71.57	15	< .001	.40	.09
French	.34	1	.559	1.14	.16
Japanese	7.19	5	.207	.33	.13
Mixture	114.26	24	< .001	.26	.07
Spanish	10.57	5	.061	.02	.15
Total	284.13	70	< .001		

Second language (L2). The effect sizes between groups (i.e., L2) were heterogeneous (see Table 4). The between-group variation was significant ($Q_{between}$ = 103.26, *df* = 9, *p* < .001). Although the total within-group variation was significant (Q_{within} = 180.88, *df* = 61, *p* < .001), most of the within-group variations were not significant (Table 3). The results suggest that L2 explained a substantial amount (103.26/284.13 = 36.3%) of the population variation of the effect size.

Table 4. *Test Results of ANOVA-Like Model on L2*

Source	*Q-Statistic*	*df*	*p*	Mean *ES*	*SE*
Between groups	103.26	9	< .001		
Within groups	180.88	61	< .001		
English	32.08	16	.010	.68	.08
eXperant	21.43	15	.124	-.09	.07
Finnish	4.99	1	.026	-.18	.35
French	35.22	8	< .001	.80	.11
Implexan[a]					
Japanese	12.54	3	.006	-.16	.22
Korean	.68	1	.408	.66	.31
Latin	2.48	3	.478	-.33	.19
Spanish	70.84	12	< .001	-.03	.10
Turkish	.61	3	.893	-.002	.18
Total	284.13	70	< .001		

[a]Only one study was in this group.

Structure. Table 5 shows that the effect sizes between groups (i.e., structure) were heterogeneous. The between-group variation was significant ($Q_{between}$ = 64.18, *df* = 5, *p* < .001). Both the total within-group variation (Q_{within} = 225.37, *df* = 65, *p* < .001) and almost all of the within-group variations were significant. The results suggest that structure only explained some (64.18/284.13 = 22.6%) of the population variation of the effect size.

Table 5. *Test Results of ANOVA-Like Model on Structure*

Source	*Q-Statistic*	*df*	*p*	Mean *ES*	*SE*
Between groups	64.18	5	< .001		
Within groups	219.95	65	< .001		
Direct O	32.27	6	< .001	.32	.13
Misc	13.34	4	.010	.57	.16
Morpheme	30.45	20	.063	-.14	.07
Sentence	65.65	19	< .001	.35	.06
Sociolin	34.31	7	< .001	.88	.14
Verb Ten	43.94	9	< .001	-.08	.12
Total	284.13	70	< .001		

Design. As we can see in Table 6, the effect sizes between groups (i.e., design) were heterogeneous. The between-group variation was significant ($Q_{between}$ = 95.33, *df* = 1, *p* < .001). Both the total within-group variation (Q_{within} = 188.80, *df* = 69, *p* < .001) and all of the within-group variations were significant (Table 5). The results suggest that design only explained some (95.33/284.13 = 33.6%) of the population variation of the effect size.

Table 6. *Test Results of ANOVA-Like Model on Design*

Source	*Q-Statistic*	*df*	*p*	Mean *ES*	*SE*
Between groups	95.33	1	< .001		
Within groups	188.80	69	< .001		
Intact	109.54	31	< .001	.65	.06
Random	79.26	38	< .001	-.11	.05
Total	284.13	70	< .001		

Mode of treatment. The effect sizes between groups (i.e., mode of treatment) were heterogeneous (see Table 7). The between-group variation was significant ($Q_{between}$ = 32.10, *df* = 2, *p* < .001). Both the total within-group variation (Q_{within} = 252.03, *df* = 68, *p* < .001) and all but one within-group variations were significant (Table 7). The results suggest that mode of treatment only explained some (32.10/284.13 = 11.3%) of the population variation of the effect size.

Table 7. *Test Results of ANOVA-Like Model on Mode of Treatment*

Source	*Q-Statistic*	*df*	*p*	Mean *ES*	*SE*
Between groups	32.10	2	< .001		
Within groups	252.03	68	< .001		
Computer	57.47	26	< .001	-.02	.06
Oral & W	187.37	37	< .001	.43	.06
Unknown	7.19	5	.207	.33	.13
Total	284.13	70	< .001		

Discussion

The results of this analysis have revealed a number of considerations that may contribute to the effect of explicit feedback on second language acquisition. As Table 2 indicates, there is significant difference between studies based on the setting in which the study took place. The studies with primary, junior high and high school settings each came from a single publication. The primary and high school settings had only 1 degree of freedom, which explains their insignificance. Although the studies in the junior high setting showed significance and a very high effect size (0.88, *p* < .001), all of them coming from a single study may contribute to the outcome. The two characteristics that are perhaps the most surprisingly different are the "College" and the "Lab" settings. The participants in both of these categories are college students, but the "College" category are studies where treatment took place in a regular class session, as opposed to the "Lab" category, where the treatment took place in a laboratory setting. As can be seen in Table 2, the classroom setting effect size was small at 0.27 and the laboratory setting was 0.01, showing virtually no effect. These results suggest that participants may have benefitted from the classroom environment more than in a one-on-one laboratory setting.

In terms of first language, French had the highest effect size of 1.14, which are also the same studies that took place in the junior high classroom, all coming from the same publication. The significance of this is related to the other characteristics which are discussed the following paragraphs. Likewise, the studies with the lowest effect, where the first language is Dutch, come from the same publication. It is possible that there are other effects from the research. For example, the language being studied in this publication was an artificial language, eXperanto. The first languages of the participants of the majority of the studies were English or a mixture, which produced a small to moderate effect. These results have likely come from the fact that the data are from multiple publications, while the Japanese studies came from a single publication and the Spanish studies from two publications.

The results of the ANOVA analysis of the second language of the learners is similar to that of the L1, where many of the studies all came from the same publication. This limitation may contribute to the range of effect sizes from -0.33 to -.88. The highest effect size also comes from the studies that took place in the junior high school setting, which is discussed in more detail in the following paragraphs. The studies with the lowest effect size (Latin) also all come from the same publication, as do the studies with the other L2s, except for Spanish and English. It is difficult to determine why there is such a difference between the effect sizes of Spanish (-0.03) and English (0.68), but one possibility is that most of the English studies were in English as a Second Language classes, where the participants were learning English in a country where English is the dominant language spoken in society. To that end, it would have been virtually impossible to control for exposure to English that the participants had outside of the study. For example, English learners may abandon their first language and speak only English at home (see Fillmore, 2000). They are also likely to speak English outside of the classroom in daily interactions with their friends in after-school activities. Language learners living in an English-speaking society also have unlimited access to television and music in English. These are all variables that a researcher cannot control when the subjects of a study are English learners living in a predominantly English-speaking society.

The target structure for the studies is the one characteristic that has the potential to have the biggest impact on the effectiveness of explicit feedback, but once again there is a significant span in the effect sizes, from -0.14 to 0.88. The highest effect comes from the junior high school studies once again. One possible reason for these studies to repeatedly have the highest effect

size is because of the target structure. These studies were the only ones to not focus on grammar, but rather on sociolinguistic rules of the language, namely the use formal versus informal expressions and register in French. It is also worth noting that in addition to the studies coming from the same publication, the participants were in a French immersion program in Canada, where the participants may have already had exposure to these sociolinguistic rules or had access to additional use of the language outside of the classroom, given that the study took place over 63 days. Verb tense had the lowest effect size, while the use of direct objects and their placement, and other forms of sentence structure (e.g., adverb placement), and other miscellaneous grammatical structures (e.g., the use of articles and participal adjectives) all had moderate effect sizes. These results may suggest that explicit feedback may be effective at learning some structures of language (e.g., sociolinguistic rules, adverb placement, etc.), but not others, such as verb tense.

When examining the design of the study there is a significant difference in the effect size of studies that used intact groups (0.65) and studies that used randomization (-0.11). These results are probably not too surprising, given that when using intact groups it is difficult to control for possible influences on the outcome, such as the ability of the learners. What is more, in many instances, the intact groups were taught by their regular classroom teachers, and most of the random studies were completed in a laboratory. Furthermore, most of the laboratory studies were done with a computer, where the intact groups were given their treatments by a live person, in the context of a regular classroom setting. This is further corroborated by the final ANOVA analysis seen in Table 7, where studies completed via computer had a negative effect size of -0.02 and human delivery through oral and written activities had an effect of 0.43. The "unknown" studies we do know were not carried out by computer, so in effect we can also say that they further substantiate the idea that feedback provided by a human has much more impact than that provided by a machine.

Conclusion

This meta-analysis has revealed a number of interesting findings that support arguments made in the body of literature examining the effect of explicit error correction on language learning.

Considerations for Research on Explicit Feedback

First, there is confusion as to what constitutes explicit error correction in studies. While there is a general consensus of what explicit error correction is, in many instances, researchers and instructors may choose to add metalinguistic elements to the feedback in order to help the learner understand why what they said was wrong. In other cases, no explanation is offered to the learner. Additionally, sometimes explicit feedback provides the correction for the learner, and sometimes there is an attempt to elicit the correction from the learner. Many of the studies included in this analysis did not describe exactly how the explicit feedback was provided, whether care was taken to control for certain extra elements (e.g., providing the correct response). This inconsistency could contribute to the variance noted in the effect sizes of the studies, but without better control and reporting in the studies, it is difficult to determine for sure if the way explicit feedback is delivered has a bigger effect on language learning.

Second, there is confusion as to whether it is acquisition that is taking place as a result of explicit feedback, the simple raising of awareness, or perhaps some other type of effect taking place as a response to the feedback. The positive effect sizes seen in the studies included in this analysis may only be an immediate response to the treatment (i.e., uptake; see Lyster & Ranta, 1997; Panova & Lyster, 2002). This also raises questions as to whether explicit feedback is long lasting. Mackey and Philp (1998) and White (1991), for example, found that corrective feedback has a positive effect in the short-term, but not the long term. In our ANOVA analyses, the length of time between pre-test and delayed post-test did not show significance. Although the results of our study show that explicit correction is not affected by length of time, a common problem when studies take place in a classroom setting, of course, is that one cannot control for the learning, practice and exposure to language as positive evidence that may take place between treatment and the post- or delayed post-test. Although White (1991) found a year later that learners had not recalled the structure learned with corrective feedback, Long (2007) suggests that researchers should develop longitudinal studies to purposefully examine the long-term effects of feedback on language acquisition, along with which one could examine the effects of the number of treatments to which a learner is exposed along with the effects of time.

A third point that is often ignored, but which may play a significant role in the effect that feedback has on language learning or acquisition is the level of “readiness” of the participant (Mackey & Philp, 1998; Mackey, Gass & McDonough, 2000). Findings in these studies led to the important

discussion about factors that may affect whether the learner is able to recall explicit correction later. Factors that fall under the "readiness" category include such characteristics as age and linguistic ability. Unfortunately, despite these studies' recommendations, most research still does not report on effects of age and linguistic level as they relate to the effect of corrective feedback. Given that many of the studies included in this meta-analysis failed to report the ages of the participants or the linguistic level of the learners, we were unable to include these two categories as factors in our final analysis. Most studies reported at a minimum the setting, some having taken place in a high school, junior high or primary school setting, while the majority were in university classes of mixed age groups. This could have also contributed to the vast span of the effect sizes across the studies in our analysis (see Table 2).

Finally, our analysis has revealed that because there are so many characteristics that appear to contribute to the effectiveness of explicit feedback, there is simply a great need for additional studies that control for these characteristics. Many of the factors we examined had very low degrees of freedom (see Tables 2-7). As we have discussed in previous sections, examples of such factors include age, linguistic ability, number of treatments, particular linguistic structures, the type of treatment used, among many others. We urge researchers in SLA to carefully design their studies in such a way as to help develop a better understanding of the very complex nature of corrective feedback.

Implications for the Language Classroom

This meta-analysis has brought forth several interesting implications for the L2 classroom. As our analysis revealed, the overall effect size of explicit error correction on second language learning is 0.27 unweighted and 0.25 weighted, which suggests a small effect, one that is often only observable through careful study (Cohen, 1988). Although our discussion has noted the difficulty in claiming that explicit feedback promotes acquisition, this analysis, which supports the findings of many of the other studies presented here, suggests that explicit feedback may raise language learners' level of awareness or noticing, at least in some circumstances. It is also possible that explicit feedback may have a role similar to that of focused practice activities, where, as VanPatten (2002) explained, such forms of instruction merely have an influence on output, but not on the developmental system. Further study is needed in order to fully understand what role explicit corrective feedback has on L2 learning and acquisition.

As we have discussed, there are still far too many questions that remain unanswered to make any conclusive comments about the effectiveness of explicit feedback, at least on long-term learning or acquisition of a second language. However, there are a number of measures that language teachers can take to enhance instruction, despite the uncertainty of whether such measures lead to acquisition or just surface level language learning. There is evidence in the body of literature that language learners require exposure to examples of correct language use (i.e., positive evidence), as well as opportunities to produce language in an interactive setting, where learners are able to repair their language based on the interactions they have with teachers and peers (see Swain, 1985, 1989; White, 1991). Most explicit feedback, however, does not provide many opportunities to repair language, which is likely to explain, at least in part, the low overall effect size of explicit correction in our analysis.

So what does this mean for the classroom? Language teachers must be committed to ensuring that language learners are not only exposed to the L2, but that learners are provided with many opportunities to interact in the L2 and to correct mistakes and errors. There is significant support in the literature for both the provision of high frequency of exposure to positive evidence, as well as language production (Bley-Vroman, 2002; Dekeyser, 2001; Ellis, 2002). The analysis provided here suggests that while explicit feedback may enhance instruction, it is important to ensure that along with feedback, instructors purposefully monitor learners to ensure that they notice their errors, have opportunities to self-correct, and that feedback be implemented with other instructional techniques (see, for example, Brandi, 2007; Omaggio Hadley, 2000).

Published 27th January 2016

Biodata

Paul Chamness Miller, Ph.D. is Professor of International Liberal Arts in the English for Academic Purposes program of Akita International University in Japan. His research interests include second language teaching practices, as well as critically examining the issues and needs of under-represented learners and teachers in primary and secondary schools.

Wei Pan, Ph.D. is Associate Professor and Director of the Research Design and Statistics Core in the School of Nursing at Duke University, USA. His research work focuses on causal inference and propensity score analysis; multilevel, structural, and longitudinal modeling; meta-analysis; psychometrics; and their applications in the educational and health sciences.

References[1]

*Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning and teaching* (Technical Report No. 9, pp. 259-302). Honolulu: University of Hawaii.

American Council on the Teaching of Foreign Languages (ACTFL). (1996). *Standards for foreign language learning: Preparing for the 21st century.* Yonkers, NY: ACTFL. Retrieved from http://www.actfl.org/files/public/execsumm.pdf.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*, 257-278.

Bley-Vroman, R. (2002). Frequency in production, comprehension, and acquisition. *Studies in Second Language Acquisition, 24*(2), 209-213.

Brandi, K. (2007). *Communicative language teaching in action: Putting principles to work.* New York: Prentice Hall.

Carlson, K. D., & Schmidt, F. L. (1999). Impact of experimental design on effect size: Findings from the research literature on training. *Journal of Applied Psychology, 84*, 851-862.

Carroll, S. (1999). Putting “input” in its proper place. *Second Language Research, 15*, 337-388.

*Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition, 15*, 357-386.

Chomsky, N. (1959). A review of B. F. Skinner's verbal behavior. *Language*, 35(1), 26-58.

[1] Asterisks (*) indicate studies included in the meta-analysis calculations.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

DeKeyser, R. (1993). The effects of error correction on grammar knowledge and oral proficiency. *The Modern Language Journal, 77*, 501-514.

*DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 17*, 379-410.

*DeKeyser, R. M., & Sokalski, K. J. (1996). The differential role of comprehension and production practice. *Language Learning, 46,* 613-642.

Diab, R. L. (2005). EFL university students' preferences for error correction and teacher feedback on writing. *TESL Reporter, 31*(1), 27-51.

*Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition, 13*, 431-469.

*De Graaff, R. (1997). The Experanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition, 19*, 249-276.

Ellis, N. C. (2002). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition, 24*, 297-339.

Fillmore, L. W. (2000). Loss of family languages: Should educators be concerned? *Theory Into Practice, 39*(4), 203-210.

Greenhouse, J. B, & Iyenger, S. (1994). Sensitivity analysis and diagnostics. In H. Cooper, & L. V. Hedges (Eds.). *The handbook of research synthesis* (pp. 383-398). New York: Russell Sage Foundation.

Hedges, L. V, & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.

Hedges, L. V. (1994). Fixed effects models. In H. Cooper, & L. V. Hedges (Eds.). *The handbook of research synthesis* (pp. 285-299). New York: Russell Sage Foundation.

Hymes, Dell H. (1966). Two types of linguistic relativity. In W. Bright (Ed.), *Sociolinguistics* (pp. 114-158). The Hague: Mouton.

Hymes, Dell H. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269-293). Harmondsworth: Penguin.

Ji, X. (2015). Error correction in college EFL writing instruction: Students' expectations and correction effects. *The Journal of Asia TEFL, 12*(1), 117-140.

*Kang, H.-S. (2009). The relative efficacy of explicit and implicit feedback in the learning of a

less-commonly-taught foreign language. *International Review of Applied Linguistics, 47*, 303-324.

*Kubota, M. (1994). The role of negative feedback on the acquisition of the English dative alternation by Japanese college students of EFL. *Institute for Research in Language Teaching Bulletin, 8*, 1-36.

Lee, J. F., & VanPatten, B. (2003). *Making communicative language teaching happen* (2nd edition). New York: McGraw-Hill.

*Leow, R. P. (1998). The effects of amount and type of exposure on adult learners' L2 development in SLA. *Modern Language Journal, 82*, 49-68.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

Long, M. H. (2007). *Problems in SLA*. New York: Lawrence Erlbaum.

*Lyddon, P. A. (2011). The efficacy of corrective feedback and textual enhancement in promoting the acquisition of grammatical redundancies. *Modern Language Journal, 95(Supplement)*, 104-129.

*Lyster, R. (1994). The effect of functional-analytic teaching on aspects of French immersion students' sociolinguistic competence. *Applied Linguistics, 15*, 263-287.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19,* 37–66.

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal, 82,* 338–356.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition, 22,* 471–497.

*Master, P. (1994). The effect of systematic instruction on learning the English article system. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 229-252). Cambridge: Cambridge University Press.

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (n.d.). Chapter 4: Foreign language activities. Tokyo: Author. Retrieved from http://www.mext.go.jp/component/english/__icsFiles/afieldfile/2011/03/17/1303755_011.pdf

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105-125.

*Nagata, N. (1995). An effective application of natural language processing in second language instruction. *CALICO Journal, 13*(1), 47-67.

*Nagata, N. (1997a). The effectiveness of computer-assisted metalinguistic instruction: A case study in Japanese. *Foreign Language Annals, 30*, 187-200.

*Nagata, N. (1997b). An experimental comparison of deductive and inductive feedback generated by a simple parser. *System, 25*, 515-534.

Omaggio Hadley, A. (2000). *Teaching language in context*. New York: Heinle & Heinle.

Papangkorn, P. (2015). SSRUIC students' attitude and preference toward error corrections. *Procedia: Social and Behavioral Sciences, 197*, 1841-1846.

Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly, 36*, 573-595.

Richards, J. C. (2006). *Communicative language teaching today*. Cambridge: Cambridge University Press.

*Salaberry, M. R. (1997). The role of input and output practice in second language acquisition. *Canadian Modern Language Review, 53*, 422-451.

Salteh, M. A., & Sadeghi, K. (2015). Teachers' and students' attitudes toward error correction in L2 writing. *Journal of Asia TEFL, 12*(3), 1-31.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.

*Staford, C. A., Bowden, H. W., & Sanz, C. (2012). Optimizing language instruction: Matters of explicitness practice, and cue learning. *Language Learning, 62*, 741-768.

Swain, M. (1985). Communicative competence: Some rules of comprehensible input and comprehensible output in its development. In S. Gass, & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Swain, M. (1989). Canadian immersion and adult L2 teaching: What's the connection? *Modern Language Learning, 73*, 150-159.

Truscott, J. (1999). The case for ''the case for grammar correction in L2 writing classes'': A response to Ferris. *Journal of Second Language Writing, 8*, 111–122.

VanPatten, B. (2002). *From input to output: A teacher's guide to second language acquisition.* New York: McGraw-Hill.

*VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition, 15*, 225-243.

*VanPatten, B., & Oikkenon, S. (1996). Explanation versus structured input in processing

instruction. *Studies in Second Language Acquisition, 18,* 495-510.

White, L. (1991). Adverb placement in second language acquisition: Some positive and negative evidence in the classroom. *Second Language Research, 7,* 122-161.

*White, L., Spada, N., Lightbown, P. M., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics, 12*, 416-432.

*Williams, J., & Evans, J. (1998). What kind of focus and on which forms? In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 139-155). Cambridge: Cambridge University Press.

*Yilmaz, Y. (2012). The relative effects of explicit correction and recasts on two target structures via two communication modes. *Language Learning, 62*, 1134-116.