Check for updates

Article

LANGUAGE TEACHING RESEARCH

Language Teaching Research
1–26
© The Author(s) 2018
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1362168818776671
journals.sagepub.com/home/ltr
SAGE

Thirty-five years of ISLA on form-focused instruction: A meta-analysis

Eun Young Kang
Lehman College, City University of New York, USA

Sarah Sok
University of California, Irvine, USA

ZhaoHong Han
Teachers College, Columbia University, USA

Abstract

This meta-analysis offers a snapshot of thirty-five years (1980–2015) of research on instructed second language acquisition (ISLA). Fifty-four empirical studies involving a total of 5,051 second language learners – sampled from six applied linguistics journals, *Applied Linguistics, Language Learning, Language Teaching Research, The Modern Language Journal, Studies in Second Language Acquisition*, and *TESOL Quarterly* – were aggregated for the effects of second language (L2) instruction, yielding an overall large effect size, g = 1.06, 95 % CI = 0.84−1.29. Data were further analysed to identify factors that can modulate the efficacy of instruction. While a minor difference was detected between explicit and implicit instruction, statistically significant effects were found for modes of outcome measures, learners' onset L2 proficiency, research settings, and intensity of instruction.

Keywords

explicit instruction, factors influencing efficacy of instruction, implicit instruction, L2 instructional efficacy, meta-analysis

Corresponding author:
Eun Young Kang, Lehman College, City University of New York, 250 Bedford Park Boulevard West, New York, NY 10468, USA.
Email: eunyoung.kang@lehman.cuny.edu

I Introduction

The role of instruction in second language (L2) development is a topic that has consistently garnered attention in applied linguistics, notwithstanding swings of the theoretical and the methodological pendulum. Early research was predominantly concerned with the influence of instruction on the rate and ultimate attainment of L2 learning (Long, 1983). Since the 1990s, however, attention has shifted to types of instruction and their relative effects (Goo, Granena, Yilmaz, & Novella, 2015; Norris & Ortega, 2000). This largely followed from the fact that by the 1990s, there had been substantial accumulation of relevant empirical studies (Goo et al., 2015). A quick overview of the shifting landscape of research on instructed second language acquisition (ISLA) can be found Nassaji (2016).

Attempts have been ongoing to take stock of findings from the ever-growing number of empirical studies (e.g. Doughty & Williams, 1998), most of them taking the form of narrative reviews. Meta-analytic procedures first debuted in Norris and Ortega (2000).

The Norris and Ortega meta-analysis employed a lockstep set of procedures – from sampling empirical studies from the literature to calculating the effect sizes of individual studies and aggregating the values into an average overall effect size – and uncovered a large effect size ($d = 0.96$), indicating that L2 instruction is strongly effective. (According to Cohen's benchmarks, 0.2 = small, 0.5 = medium, and 0.8 = large effect.)

Over the years, this work has inspired other scholars to attempt similar meta-syntheses of research on instructional efficacy. The present study was yet another such attempt, but with an eye toward updating and extending the results from Norris and Ortega (2000). Our data sample comprised primary studies published in six journals, *Applied Linguistics, Language Learning, Language Teaching Research, The Modern Language Journal, Studies in Second Language Acquisition*, and *TESOL Quarterly*, spanning 35 years (1980–2015). Data analysis pursued the same two-pronged focus that had guided the Norris and Ortega study: the overall efficacy of instruction and the effects of modulating factors.

In the sections that follow, we first briefly discuss meta-analysis as a way to synthesize primary research findings on the role of instruction in second language acquisition (SLA). We then summarize and review the findings from previous meta-analyses. After that, we report on our study and discuss the main findings both in their own right and in relation to Norris and Ortega (2000).

II Meta-analysis of effects of instruction

Meta-analysis as a tool of research synthesis is a newcomer in SLA research. Since Norris and Ortega (2000), it has quickly gained recognition and momentum (Ellis, 2015). To date, meta-analysis has been applied to an array of topics, including, but not limited to, grammar, pragmatics, and pronunciation (Plonsky, 2016).

Where meta-analysis of the general effects of instruction is concerned, Norris and Ortega (2000) provided a critical baseline. Employing a stringent set of screening criteria, the study sampled 45 studies on L2 grammar instruction published between 1980 and 1998 in a variety of outlets, comparing and aggregating their effects relative to the condition of learning without instruction (i.e. meaning-oriented communication), and reporting a robust average effect size (*d* = 0.96). In addition, explicit instruction appeared to result in greater learning gains than implicit instruction (*d* = 1.13 vs. *d* = 0.54). Further analyses revealed that outcome measures requiring learners' spontaneous use of the L2 (e.g. free production) yielded lower effect sizes than those requiring controlled use of the L2 (e.g. multiple choice and cloze). Moreover, duration of instruction seemed to impact the efficacy of instruction, with short-term instruction having larger effects on learning than longer-term instruction. Still, Norris and Ortega were quick to caution potential confounding factors like the intensity of instruction and the types of linguistic structures targeted, calling for further investigation.

Ten years later, Spada and Tomita (2010) meta-analysed the effects of three variables: type of instruction (implicit/explicit), type of target linguistic features (simple/complex), and type of outcome measures (controlled/free production). Their data sample comprised 30 empirical studies published between 1990 and 2006, yielding findings that largely paralleled Norris and Ortega's. First, regardless of the type of linguistic target, explicit instruction was found more effective than implicit instruction on both simple (*d* = 0.73) and complex (*d* = 0.88) linguistic features (simple, *d* = 0.33; complex, *d* = 0.39). Second, following explicit instruction, learners performed better on both controlled and free production measures, contra the authors' own expectations and Norris and Ortega's finding. Spada and Tomita were, however, wary of possible overestimation of the efficacy of explicit instruction, driven by the skewed distribution of studies in their data sample – greater representation of explicit instruction than implicit instruction. Another data-related bias was that most of the studies adopted outcome measures that involved controlled use of L2 knowledge in isolation, which may have played up the effects of explicit instruction. The authors called for further investigation of the effects of implicit and explicit instruction, including the types of knowledge developed.

Meta-analytic attempts to address these concerns have surfaced in recent years. An example is a meta-analysis by Goo, Granena, Yilmaz, and Novella (2015) exploring the relative effects of types of instruction and moderating factors. The data sample consisted of 34 studies, 11 of them overlapping with the data sample of Norris and Ortega (2000). The results once again confirmed the overall positive effects of L2 instruction (*g* = 1.03) and the advantage of explicit (*g* = 1.29) over implicit instruction (*g* = 0.77).[1] But Goo et al. found a higher overall mean effect size for implicit instruction than reported by Norris and Ortega (2000) and Spada and Tomita (2010). The authors attributed the difference to the inclusion of 21 new studies but, essentially, to the surge of interest in investigating implicit instruction. Consistent with Norris and Ortega (2000), Goo et al. found, as well, that regardless of the type of instruction, selected or controlled response measures tended to yield a larger mean effect size than free production measures.

In spite of differences in the timing and sample size, the three meta-analyses all have converged on the finding that the effects of L2 instruction are differential: by type of instruction (e.g. explicit vs. implicit) and type of outcome measure (controlled vs. free production). Additional mitigating variables on L2 instructional efficacy have been uncovered, among them the mode of instruction (computer-mediated vs. face-to-face), the research setting (labs vs. classrooms; second language vs. foreign language), proficiency, and the educational context (see, for example, Mackey & Goo, 2007; Li, 2010; Russell & Spada, 2006).

The present meta-analysis sought to substantiate these findings by both expanding the data sample and extending the timeline (1980–2015) and by employing a newer meta-analytic approach than the one taken by Norris and Ortega (2000). In lieu of averaging out each effect size of the studies, a random-effects statistical model was adopted for ascertaining the overall effect of instruction. The random-effects model was deemed appropriate to aggregate the effect sizes of the included studies due to the methodological heterogeneity observed among them (see, for example, Li, 2010; Shintani, Li, & Ellis, 2013). Additionally, instead of comparing combined effect sizes for individual variables, *Q between* (Q*b*) tests were conducted to assess the magnitude of the impact of moderator variables. This type of tests has become increasingly common for moderator analyses in the field of SLA (see, for example, Kang & Han, 2015; Li, 2010; Shintani et.al., 2013). Finally, the present study investigated seven additional moderator variables: delivery and mode of instruction, linguistic targets, L2 proficiency, educational context of L2 learners, and research settings (see Table 4 below).

The questions guiding the present meta-analysis were, as in Norris and Ortega (2000), two:

1. What is the overall effectiveness of instruction?
2. What variables effectively moderate the effect of instruction?

III Method

1 Literature search

Data sampling began with drawing up a list of search terms, resulting in the following selection: *focus on form(s), form-focused instruction, negative feedback, grammar instruction, explicit or implicit instruction, error correction, written feedback, recasts, prompts, models, textual enhancement, input enhancement*, and *processing instruction*. These items were then applied to a search of relevant primary studies published between 1980 and 2015 in *Applied Linguistics, Language Learning, Language Teaching Research, The Modern Language Journal, Studies in Second Language Acquisition*, and *TESOL Quarterly*. Admittedly limited in the scope of sampling, the six journals were purposely chosen to enable a comparison with the results from Norris and Ortega (2000) where the six journals provided the staple of their sample. The abstracts of the studies were machine scanned, followed by a manual review. The reference section of Norris and Ortega (2000) was crosschecked for potentially missing studies published before 2000.

2 Eligibility criteria

The screening criteria were adapted from Norris and Ortega (2000). First, the studies were interventional in nature and targeted morphosyntactic elements. Second, the studies had an experimental or quasi-experimental design, including a comparison or a control group. Third, the studies measured L2 learning outcomes vis-à-vis the focus of instruction. Fourth, the instruction was conducted on L2 learners.[2] Above all, the studies provided sufficient statistical data for the estimation of mean differences between the treatment and control conditions.

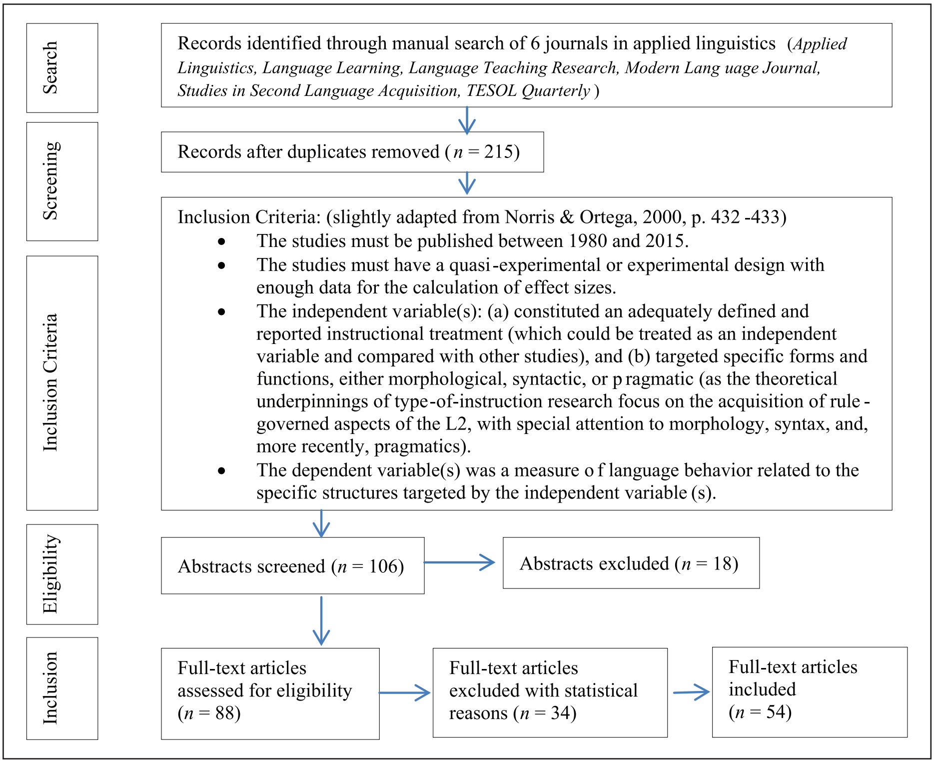

Figure 1. Data sampling.

3 Reliability

Two researchers independently reviewed all titles and abstracts. The initial search generated 215 potential studies (see Sok, Kang, & Han, this issue). Figure 1 illustrates the process of data sampling. The studies that fully met the selection criteria were then double checked independently by the two researchers. Any disagreements were resolved through discussion. A consensus was reached for a total of 54 studies,[3] of which 15 studies overlapped with Norris and Ortega's (2000) sample and the remaining 39 studies were new: published between 2000 and 2015.

4 Coding

Based on Norris and Ortega (2000), a coding scheme was created, with three categories: (1) type of instruction, (2) type of outcome measure, and (3) methodological feature (see Table 1).

a Type of L2 instruction. In Norris and Ortega (2000), type of instruction was coded in terms of (1) explicit vs. implicit instruction and (2) focus on forms (FonFS) vs. focus on form (FonF) vs. focus on meaning (FonM). The two superordinate variables, explicitness (implicit vs. explicit) and attention to form (FonFS vs. FonF) intersected to yield four

Table 1. Coding scheme.

Main category		Variables
Second language (L2) instruction	[Type]	Explicit vs. implicit
	[Delivery]	Computer-mediated vs. face-to-face
Outcome measures	[Type]	Metalinguistic judgment vs. selected response vs. constrained constructed response vs. free constructed response
	[Mode]	Written vs. oral vs. combined
Methodological features		Linguistic target (Morphemes vs. syntax vs. pragmatics)
		L2 proficiency (High vs. Mid vs. Low)
		Educational setting (Foreign Language vs. Second Language)
		Educational context (Elementary vs. Secondary vs. University vs. Language Institute)
		Research setting (Classroom vs. Lab)
		Duration of instruction (in days)
		Intensity of instruction (in hours)

instructional conditions: FonF explicit, FonF implicit, FonFS explicit, and FonFS implicit. In the present study, however, care was taken to mitigate a typical weakness of previous research as a result of conducting moderator analysis with an uneven or low number of studies (Han, 2015). We employed a dichotomous coding of the implicit versus the explicit, as in Goo, Granena, Yilmaz, & Novella (2015).

Following Norris and Ortega (2000), instruction was coded as explicit if it involved 'rule explanation' or 'if learners were directly asked to attend to particular forms and to try to arrive at metalinguistic generalizations on their own,' but as implicit if 'neither rule presentation nor directions to attend to particular forms were part of a treatment' (p. 437). But going beyond Norris and Ortega (2000), and inspired by Ziegler (2016), we added delivery of instruction (computer mediated vs. face to face) to our coding scheme to both provide a contextual dimension to our meta-analysis and capture a growing number of primary studies on the effects of computer-mediated instruction.

b Type of outcome measure. Outcome measures were coded for the mode (written vs. oral vs. combined) and, as in Norris and Ortega (2000), for the type of response required (metalinguistic judgment vs. selected response vs. constrained constructed response vs. free constructed response). Learners determine the grammaticality of sentences on metalinguistic judgments; choose a correct answer from a range of options on selected responses; produce instructionally targeted linguistic elements in a limited context on constrained constructed responses; and demonstrate their ability to use the instructional targets spontaneously on free constructed responses.

c Methodological feature. It breaks down into linguistic target, participants' age, L2 proficiency, participant's first language (L1), and educational setting. These were coded strictly according to what was originally reported in the primary studies, except for the variable of L2 proficiency, which typically was underreported and inconsistently measured across primary studies. Inferences were made in this case, as did Norris and Ortega (2000): up to two semesters of instructed learning was coded as low proficiency, three to four semesters as mid proficiency, and five or more semesters as high proficiency.

New to the coding in the present study was also a set of treatment related variables: duration of instruction, intensity of instruction, and intensity of treatment sessions. The duration of instruction was coded in days, while intensity was coded in hours for the entire intervention and in minutes for individual treatment sessions.

Each variable in the coding scheme was simultaneously considered a potential moderator contributing to variance in effect size across the studies. The results of coding were recorded in an Excel spreadsheet. Two researchers independently coded 60% of the studies, and together, they double coded a random sample of 10% of the studies. The inter-rater agreement was 97.7%, and any discrepancies were resolved via consensus.

5 Analysis

All analyses were conducted using the Comprehensive Meta-Analysis software (CMA, version 2.0; Borenstein, Hedges, Higgins, & Rothstein, 2005). Given that the primary studies reported effects of different modes of L2 instruction with different scales using heterogeneous outcome measures, our first step was to transform the results of each outcome measure into comparable units called Hedges' *g*. Although Cohen's *d* is commonly adopted in L2 meta-research, it is allegedly biased upward for studies with small samples ($n < 20$) (Lipsey & Wilson, 2001). To mitigate the bias, in the present analysis Hedges' *g*, calculated on the basis of posttest means, standard deviations, and sample sizes of treatment and control (or comparison) groups, was adopted as an alternative index of effect size (compare Goo, Granena, Yilmaz, & Novella, 2015). Where such indexes were absent in the data sample, a *t* value for posttest mean differences between the L2 treatment and the control (or comparison) groups was used to compute Hedges' *g*.

A random-effects model was used to estimate the overall effect sizes and 95% confidence intervals. Unlike a fix-effects model, it tolerates variations in intervention procedures, participants, and study design (Borenstein, Hedges, Higgins, & Rothstein, 2011; Shintani, Li, & Ellis, 2013). Each primary study was taken as contributing a single effect size towards the overall effect of instruction. Such practice has been recommended by meta-analysts in that having a single independent estimate of the effect per study – even if the study employs multiple outcome measures – is crucial to the validity of the meta-analytic results (Borenstein et al., 2011). Similarly, when studies included both types of instruction, the effect sizes of implicit and explicit instruction were averaged out within each study to arrive at one effect size. Likewise, effect sizes were averaged for inclusion in the moderator analysis of research setting (classroom vs. lab).

In addition, to mitigate issues of dependence arising in studies reporting more than one experiment (e.g. DeKeyser & Sokalski, 1996), we aggregated the results, as suggested in Borenstein et al. (2011).

Once the overall effect size was computed, individual effect sizes were examined, using the *Q* test to determine whether variability exceeded the limit licensed by sampling

Table 2. Sources and number of primary studies.

Journal	*n*	Percentage
Studies in Second Language Acquisition (SSLA)	21	38.8
Language Learning (LL)	12	22.2
The Modern Language Journal (MLJ)	9	16.7
Language Teaching Research (LTR)	6	11.1
TESOL Quarterly (TQ)	3	5.6
Applied Linguistics (AL)	3	5.6
Total	54	100

errors (Lipsey & Wilson, 2001). A series of moderator analyses were conducted using a random-effects model to identify the sources of variation. However, if for a given moderator variable there were less than five subsamples ($k < 5$), a fixed-effects model was adopted, as recommended by Borenstein et al. (2011). *Q between* (Q*b*) tests were used to identify potential moderator variables, and the Q*b* value reaching significance was taken as an indication of meaningful contribution to the general effect size.

For each moderator analysis, the 'one study, one effect size' principle (Borenstein et al., 2011; Scammacca, Roberts, & Stuebing, 2014) was upheld in order to circumvent the issues of statistical dependence noted earlier. This resulted in the exclusion of, for example, Goo's (2012) study from the moderator analysis for type of instruction. The study was thematically on corrective feedback and included two treatment groups—participants received either recasts or metalinguistic information—and one control group, with the two types of treatment respectively operationalizing implicit and explicit instruction and seemingly contributing two effect sizes. But the study tapped the same control group for comparing and isolating the effects of the two types of treatment. The results for each type of treatment were, therefore, deemed non-independent.

The strict execution of the principle meant also the preclusion of type of outcome measure as a moderator in the analysis. Seventy-two percent of the primary studies in the present data sample employed more than one measure, often a combination of metalinguistic judgment, selected response, constrained constructed response, and/or free production. According to Borenstein et al. (2011), only one dependent variable from each study should be extracted in a moderator analysis in order to control for dependence. Only a small number of studies in the present sample fit the bill, i.e. had a single outcome measure.

IV Results

1 Descriptive overview of the data sample

As shown in Table 2, 38.8% ($n = 21$) of the primary studies came from *Studies in Second Language Acquisition*, 22.2% ($n = 12$) from *Language Learning*, 16.7% ($n = 9$) from *The Modern Language Journal*, 11.1% ($n = 6$) from *Language Teaching Research*, 5.6% ($n = 3$) from *TESOL Quarterly*, and 5.6% ($n = 3$) from *Applied Linguistics*. As displayed in Figure 2, research on instructional efficacy has seen a spike since 2006; more than half of the 54 studies were published between 2006 and 2015. About three quarters of the

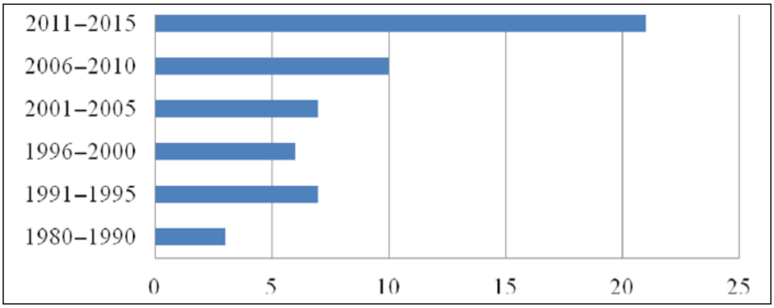

Figure 2. Studies published during 1980–2015.

studies (72. 2%, *n* = 39) were conducted in a foreign language (FL) setting, and about 27.7% in a second language (L2) setting where, by definition, the target language is spoken outside the classroom. Furthermore, 61.1 % (*n* = 33) of the studies involved face-to-face instruction, about 3.7% (*n* = 2) were delivered via pen and paper, another 27.8% (*n* = 15) via a computer, and the remaining 7.4% (*n* = 4) a blended format of face-to-face instruction and computer-mediated practice. The average span of instruction was 11.8 days, with considerable variation (*SD* = 17.4 days; minimum = 1 day; maximum = 90 days). On average, the number of instructional sessions per week was 2.2 (*SD* = 1. 5; minimum = 1; maximum = 7), each lasting 48.2 minutes (*SD* = 39.1 minutes; minimum = 10 minutes; maximum = 180 minutes). Three of the 54 studies did not provide information on the duration of instruction, 9 studies did not provide information on intensity, and 22 studies did not include information about the duration of each session.

The combined total of participants across the primary studies was 5,051, and the average number of participants of an experimental group was 29 (*SD* = 24.9; minimum = 4; maximum = 156). The majority of the participants were adults (55.6%, *n* = 31), followed by adolescents (18.5%, *n* = 10) and young learners (13%, *n* = 7). Three studies (5.6%) involved a mixed group of adults and adolescents. More than one third of the studies focused on either intermediate learners (44.4%, *n* = 22) or beginners (31.5 %, *n* = 19), five studies (9.3%) investigated more than one proficiency level, and another five studies (9.3%) advanced learners. Three studies offered little information on participants' proficiency level, and four studies did not report age. Further descriptive information on the primary studies can be found in Appendix 1 (see also Sok, Kang, & Han, this issue).

2 Overall effect of instruction

The immediate effect of instructional intervention was computed for the 54 primary studies, yielding an average large effect size estimate of $g = 1.06$ ($SE = .011$, confidence interval [CI] = 0.84–1.29) (for the effect sizes of the individual studies, see Appendix 2). The Q statistic, however, showed significant heterogeneity ($Q = 367.1$; $df = 53$; $p < .0001$), suggesting substantial idiosyncrasies across the studies. An analysis of 37 studies that reported sufficient statistical information on delayed posttests yielded an overall delayed effect size of $g = .93$ ($SE = .12$, CI = 0.68–1.16), suggesting high retention.[4]

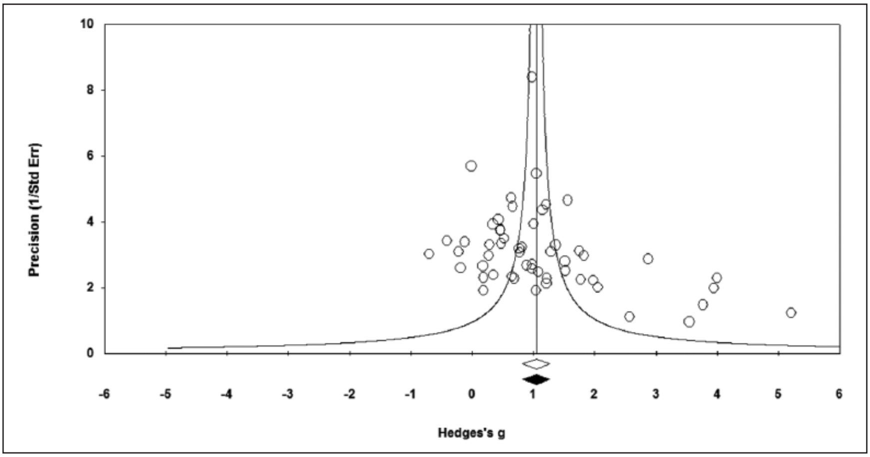

Figure 3. Funnel plot of standard error by Hedges' *g*.

3 Publication bias

A publication bias may arise when studies with significant results are overrepresented in the meta-analysis (Lipsey & Wilson, 2001). Sampling studies from journals, as was the practice in the present study, is particularly liable to such a bias. Accordingly, a series of analyses were undertaken to check whether the bias was inherent in the data sample. First, a funnel plot was created and visually inspected to check the distribution of the studies that reported statistically significant results. As shown in Figure 3, the effect sizes of the studies, represented by open circles, are distributed roughly in a symmetrical way around the mean effect size represented by an open diamond at the bottom, suggesting no substantial bias. Second, a statistical method known as the trim and fill procedure (Duval & Tweedie, 2000) was applied to estimate the number of missing studies owing to a publication bias, and to re-impute an overall effect size by incorporating the hypothetical missing studies. The filled diamond in Figure 3 represents a re-imputed value of the observed effect size, which takes into account a possible publication bias. As shown, the re-imputed and the observed overall values are not different. Results from the trim and fill analysis revealed that there were no missing studies. Finally, a fail-safe *n* was computed, the result implicating that 7,066 studies would be needed to invalidate the effects of L2 instruction found in the present meta-analysis, a value outstripping the standard benchmark whereby a fail-safe result of $5k + 10$ is deemed robust, where *k* indicates the number of primary studies included (Rosenthal, 1991). Therefore, it can be safely concluded that the research domain under investigation, i.e. instructional efficacy, is adequately represented in the current data sample.

4 Moderator analyses

The purpose of moderator analyses was to determine the extent to which the overall effectiveness of instruction was modulated by other variables. To that end, the effect sizes of subsets of studies were computed.

Table 3. Type of second language (L2) instruction as a moderator.

Moderator variable	k	Mean hedges' g (*SE*)	95% CI	$Q_{between}$
Immediate:				
Explicit	12	1.11 (0.18)	[0.74–1.47]	0.75
Implicit	18	1.38 (0.26)	[0.87–1.89]	
Delayed:				
Explicit	11	0.77 (0.14)	[0.50–1.04]	5.96*
Implicit	11	1.76 (0.34)	[1.01–2.52]	

Notes. Immediate = aggregated results of immediate posttests, delayed = aggregated results of delayed posttests, k = total number of studies, ES = effect size, CI = confidence interval, *$p < .05$. Total k is not always equal because of missing data.

Table 3 summarizes the results on type of instruction. In order to determine the short-term and long-term effects of explicit versus implicit instruction, analyses were conducted separately for immediate and delayed tests. On the short-term side, the Q statistic revealed no significant difference between explicit and implicit instruction, $Q_{between} = 0.75$, $df = 1$, $p > .05$. On the long-term side, however, a statistically significant difference was found for the delayed posttests, $Q_{between} = 5.96$, $df = 1$, $p < .05$, with implicit instruction ($g = 1.76$, $SE = 0.83$) outperforming explicit instruction ($g = 0.77$, $SE = 0.14$).

Table 4 summarizes the effects of nine additional variables, with the effect sizes from immediate posttests as the dependent variable. First, when the mode of outcome measure was analysed, the studies that used oral assessment measures ($g = 1.03$, $SE = 0.17$) or both oral and written measures (g = 1.02, $SE = 0.06$) yielded a significantly larger mean effect than studies utilizing written measures only ($g = 0.73$, $SE = 0.06$); $Q_{between} = 12.33$, $df = 2$, $p < .05$). Second, on mode of instruction, the average effect of face-to-face instruction was about the same as that provided via a computer ($g = 0.83$, $SE = 0.05$ for face-to-face vs. $g = 0.86$, $SE = 0.08$ for computer-mediated; $Q_{between} = 0.17$, $df = 1$, $p > .05$). Third, on linguistic target of instruction, the largest effect size was found for syntax ($g = 0.94$), followed, in turn, by morphology ($g = 0.85$) and pragmatics ($g = 0.32$), with significant difference ($Q_{between} = 15.36$, $df = 2$, $p < .001$). Fourth, L2 proficiency was found to be a significant moderator ($Q_{between} = 6.68$, $df = 2$, $p < .05$), with instruction having a greater effect on novice learners ($g = 1.45$, $SE = 0.25$) than on intermediate ($g = 0.70$, $SE = 0.15$) and advanced learners ($g = 0.88$, $SE = 0.14$). Fifth, educational setting was found not to be a significant moderator ($g = 0.80$, $SE = 0.07$ for the second language setting vs. $g = 0.90$, $SE = 0.05$ for the foreign language setting; $Q_{\text{between}} = 1.52$, $df = 1$, $p > .05$). Sixth, educational context was not a significant moderator ($g = 1.21$, $SE = 0.20$ for elementary schools, $g = 1.22$, $SE = 0.28$ for secondary schools, $g = 1.04$, $SE = 0.16$ for universities, $g = 0.96$, $SE = 0.41$ for language institutes; $Q_{between} = 0.77$, $df = 3$, $p > .05$). Seventh, research setting significantly affected the instructional effect of studies conducted in labs ($g = 1.06$ $SE = 0.07$) versus in classrooms ($g = 0.78$, $SE = 0.05$), $Q_{between} = 10.61$, $df = 1$, $p < .05$). Eighth, duration of instruction as a whole had a significant influence on the instructional effect of studies ($Q_{between} = 7.04$, $df = 2$, $p < .05$), though no significant difference was found between short ($g = 1.22$, $SE = 0.18$) and long ($g = 1.10$, $SE = 0.23$) treatment. Finally, intensity of instruction was a significant moderator ($Q_{between}$

Table 4. Moderator analyses of outcome measures and methodological features.

Moderator variable	*K*	Mean hedges' g (*SE*)	95% CI	$Q_{between}$
Mode of outcome measures:				
Written	22	0.73 (0.06)	[0.61–0.85]	12.33*
Oral	4	1.03 (0.17)	[0.70–1.36]	
Combined	24	1.02 (0.06)	[0.90–1.15]	
Delivery of instruction:				
Computer-mediated	15	0.86 (0.08)	[0.71–1.01]	0.17
Face-to-face	33	0.83 (0.05)	[0.73–0.93]	
Linguistic targets:				
Morpheme	35	0.85 (0.06)	[0.74–0.96]	15.36**
Syntax	15	0.94 (0.01)	[0.80–1.07]	
Pragmatics	2	0.32 (0.14)	[0.03–0.06]	
L2 proficiency:				
Low	19	1.45 (0.25)	[0.96–1.95]	6.68*
Medium	22	0.70 (0.15)	[0.41–0.99]	
High	5	0.88 (0.14)	[0.62–1.15]	
Educational setting:				
Foreign language	36	0.90 (0.05)	[0.80–1.01]	1.52
Second language	15	0.80 (0.07)	[0.66–0.93]	
Educational context:				
Elementary	7	1.21 (0.20)	[0.82–1.60]	0.77
Secondary	11	1.22 (0.28)	[0.68–1.77]	
University	27	1.04 (0.16)	[0.71–1.36]	
Language institute	6	0.96 (0.41)	[0.15–1.76]	
Research setting:				
Classroom	26	0.78 (0.05)	[0.68–0.89]	10.61*
Lab	28	1.06 (0.07)	[0.93–1.19]	
Duration of instruction:				
Short ($x < 7$ days)	33	1.22 (0.18)	[0.87–1.56]	7.04*
Medium ($8 < x < 14$ days)	8	0.59 (0.18)	[0.24–0.93]	
Long ($x > 14$ days)	10	1.10 (0.23)	[0.65–1.55]	
Intensity of instruction:[a]				
Brief (x 1hr)	13	1.81 (0.34)	[1.15–2.47]	8.04*
Short (1h < x 3h)	10	0.76 (0.19)	[0.39–1.13]	
Medium (3hr < x 6hr)	6	1.06 (0.40)	[0.29–1.83]	
Long ($x > 6$ hours)	1	0.65 (0.43)	[–0.19–1.49]	

Notes. *$p < .05$, **$p < .001$. [a]A reviewer commented that the categorization is somewhat artificial. We simply followed – for the purposes of our study, and it was necessary that we do – the scheme adopted by Norris and Ortega (2000).

= 8.04, $df = 3$, $p < .05$). Briefer instruction ($x < 1$hr) ($g = 1.81$, $SE = 0.34$) had a much larger effect on L2 learning than longer instruction ($x > 7$hr) ($g = 0.65$, $SE = 0.43$).

V Discussion

The present meta-analysis sampled 54 experimental studies from six applied linguistics journals spanning 35 years (1980–2015) and yielded a large effect size, highly analogous to that reported in Norris and Ortega (2000) and Goo, Granena, Yilmaz, and Novella (2015), thus further attesting to the efficacy of instruction.

Additional analyses of ten other variables revealed that type of instruction, above all, was a significant moderator. While both explicit ($g = 1.11$) and implicit ($g = 1.38$) instruction were found to have a large effect on L2 learning as evidenced on immediate post-tests, implicit instruction ($g = 1.76$) appeared to have a significantly longer lasting impact on learning, as revealed by delayed outcome measures, than explicit instruction ($g = 0.77$). The latter finding, consistent with Goo et al. (2015), was a major reversal of that of Norris and Ortega (2000), where implicit instruction reportedly had a moderate effect on learning ($d = 0.54$), outstripped by explicit instruction ($d = 1.13$). The discrepancy might have stemmed from the data samples used.[5] The present meta-analysis included 39 new studies which surfaced a heightened interest in implicit instruction.[6] Most of these studies were lab-based and tended to yield greater effect sizes.

Similarly, outcome measure was found to be a significant moderator in the present analysis. The measures deployed in the new studies were more diverse than in earlier studies, with greater use of spontaneous free production, than reported in Norris and Ortega (2000). Earlier studies, as Doughty (2003) and Spada and Tomita (2010) have observed, had mainly used controlled measures of explicit knowledge, thereby contributing to the reportedly larger overall effect size for explicit instruction. In contrast, nearly 41% percent of the present data sample ($k = 22$) incorporated outcome measures eliciting learners' spontaneous use of the L2 or implicit knowledge, and this might, likewise, have contributed to the greater effect size of implicit instruction.[7]

The mode of outcome measure emerged also as a significant moderator of the effectiveness of instruction. Studies employing both oral and written measures yielded larger effects than studies that included only written assessments. It seems that when more, not fewer, measures were used, there was a better chance of capturing the effects of instruction (see Goo et al., 2015). Between the written and oral modes of assessment, it appeared, with the caveat of a very small sample, that learners performed much better on oral than on written measures. However, a similar caveat should extend to a potential confound in the mode of outcome measure with the type of outcome measure. Two out of four studies that employed oral tests assumed the selected response format (e.g. multiple choice tests), whereas 12 out of 22 studies using written measures invoked metalinguistic judgment tasks. It is possible that metalinguistic judgments are harder for L2 learners than selecting responses (Norris & Ortega, 2000). Further research is warranted to clarify the tangled relationship between the mode and type of measure.

The present meta-analysis has also uncovered type of linguistic target as a significant moderator. Instruction targeting syntax yielded a significantly larger effect size than instruction on pragmatics and on morphology. Outside the domain of ISLA, research has shown that morphological forms are harder to acquire due to their low salience (Ellis, 2005) and are highly susceptible to crosslinguistic influence (Han, 2011, 2014). The relative effects of instruction on syntax versus morphology remain a target of future meta-analysis.

Proficiency, too, can significantly modulate the effects of instruction. While Norris and Ortega (2000) did not compare the average effect sizes of studies based on the level of L2 proficiency, our analysis revealed that instruction was beneficial for all levels, but particularly for novice learners, a finding standing to reason: the greater lack of L2 knowledge means that beginning learners are likely to be in greater need of external support.

The setting in which a study was conducted is another significant moderator. The effect of instruction was significantly larger for studies carried out in labs, where there are fewer distractions and therefore it is easier to engage learners' attention to the target of instruction (see, for example, Li, 2015; Mackey & Goo, 2007), than in classrooms.

Just as important, the present meta-analysis discounted several variables as having sufficient influence over effect sizes of instruction. These include amount of instruction– neither longer nor more intensive instruction had superior effects over shorter and less intensive instruction (see also Norris & Ortega, 2000); educational setting (i.e. second vs. foreign language) – the finding is inconclusive across several meta-analyses (see, for example, Goo et al., 2015; Li, 2010; Kang & Han, 2015); educational context (i.e. primary vs. secondary vs. university vs. language institute) – this is open to further investigation as well, as the aggregated results are inconclusive (see, for example, Kang & Han, 2015; Lee, Jang, & Plonsky, 2015); and last but not least, delivery of instruction (i.e. face-to-face or computer-mediated) – the finding is consistent with that of Ziegler (2016), that the efficacy of computer-based L2 interaction is not significantly different from face-to-face instruction, notwithstanding the purported virtues of the former, such as allowing learners to control the speed at which information is delivered, reinforcing their correct responses by providing instant feedback, and giving them opportunities to negotiate (Frederickson, Reed, & Clifford, 2005; Smith, 2004).

Still, as with each of the more powerful moderators, the findings on each of these less influential ones should be taken with a grain of salt. Each of these variables is likely to have conspired with some other, as yet unidentified, variables. Take the finding that shorter intervention outstrips longer intervention as an example. Tangled with this finding might have been the scope of instruction, a variable that did not make it to the list of moderators because the relevant data sample was too small. The short-term studies generally had a narrower focus, were implemented with individual students, and were conducted in lab settings, while the longer-term studies typically involved whole classes where instructional attention was more dispersed.

VI Conclusions

The present meta-analysis contributes to the existing literature in two main ways. First, it updates the findings from the Norris and Ortega (2000) study, using a more extended and current data sample. Second, it affirms several findings from previous meta-analyses, not the least that instruction is effective. Additionally, this meta-analysis explored a host of moderating variables (see Table 4), going beyond the three investigated by Norris and Ortega (2000): type of instruction, type of outcome measure and intensity of instruction.

That said, several limitations are apparent in the present study. First, the sample of primary studies came only from six journals. Though statistically it is free from a sampling bias, the study could have been done with a broader sample drawn from a variety of sources, including other journals, edited volumes and unpublished dissertations, and the findings would likely have been more nuanced. Second, as is typical of all meta-analyses, the present study might have fallen prey to artifacts emanating from use of contrived criteria for data sampling and coding. Moreover, the analysis of the relative effectiveness of explicit and implicit instruction was narrowly and grossly contained in comparing either an explicit or an implicit condition with a control group. While doing so followed a precedent set by Norris and Ortega (2000), the effectiveness of instruction could have been more accurately and adequately gauged by looking only at those studies that investigated both conditions, as in Goo et al. (2015). But then, these are methodological options, each having its strengths and weaknesses. The Goo et al. approach, for instance, arguably shed more valid light on the relative effectiveness of explicit versus implicit instruction. In allowing each primary study to contribute separate effect sizes for implicit and explicit instruction, the authors had relatively equal samples for the moderator analyses. Yet equally palpable is the downside that there was a much smaller sample of primary studies to begin with, a chief concern of any type of inferential statistical analysis. Because of the need to make such kinds of subjective decisions, the findings of any meta-analysis must perforce be interpreted as suggestive rather than definitive (Han, 2015). However, when multiple meta-analyses report similar, if not identical, findings, those findings must be taken seriously.

Conflict of Interest

The findings of this meta-analysis were based on 54 sample studies included in Sok, Kang, and Han (this issue). This study incorporated quantitative mata-analytic techniques to summarize and interpret the findings reported in the 54 studies.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

1. Goo et al. adopted Hedges' *g* as an effect size estimator as opposed to the more common Cohen's *d* as used in most other studies. Although there are some differences in how Cohen's *d* and Hedges' *g* are computed, the interpretations of these statistics are similar in that 0.2, 0.5, and 0.8 denote, respectively, small, moderate, and large effects (Lipsey & Wilson, 2001).
2. The current meta-analysis included only studies conducted with groups of L2 learners, excluding those that collected data from both L1 and L2 learners (e.g. Van Beuningen, De Jong, & Kuiken, 2012). Moreover, following Norris and Ortega (2000), studies that focused on pronunciation (e.g. Saito, 2013) or vocabulary instruction were excluded, as well.
3. A reviewer observed that several more studies could have been included in the meta-analysis. However, of the three studies s/he mentioned, only one study (Benati, 2005) met our inclusion criteria.

4. Several studies reported more than one delayed posttest administered at different time intervals, in which case the first delayed posttest was adopted for the analysis of the delayed effect. On average, there was a 32-day interval between the instructional treatment and the delayed posttests.
5. In the current dataset, some studies investigated only one type of instruction (implicit or explicit instruction), some both types, and others a particular type of L2 instruction with elements of both implicit and explicit instruction (e.g. rule explanation followed by communicative activities). In our analysis, we focused only on the first type of studies and found a larger effect size of implicit instruction. It is possible that this could have been an artifact of sampling procedures. When we analysed the second type of studies ($k = 9$), we found a larger overall effect size of explicit instruction ($g = 0.77$, $SE = 0.26$) than that of implicit instruction ($g = 0.42$, $SE = 0.13$) even though the difference was not statistically significant ($Q_{between}$ 1.37, $df = 1$, $p > .05$). When interpreting the results, it is important to keep in mind that the dataset feeding the present meta-analysis came only from six journals, without aiming at a particular type of instruction, explicit or implicit. Future meta-analysts could focus more specifically on one or the other.
6. The present study included 12 and 18 studies on explicit and implicit instruction, respectively. Norris and Ortega (2000) included 23 studies and 6 studies on explicit and implicit instruction.
7. Following Ellis (2005), metalinguistic judgment tasks like grammatical judgment tests (GJT) can be considered measures of explicit or implicit knowledge depending on whether they are timed or not. Not all studies specified how measures were employed in detail. We were not able to code all measures either as explicit or implicit knowledge clearly due to the lack of information provided in the original studies.

References

Studies included in this meta-analysis are marked with an asterisk.

*Akakura, M. (2012). Evaluating the effectiveness of explicit instruction on implicit and explicit L2 knowledge. *Language Teaching Research*, *16*, 9–37.

*Ammar, A., & Spada, N. (2006). One size fits all?: Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition*, *28*, 543–574.

*Andringa, S., de Glopper, K., & Hacquebord, H. (2011). Effect of explicit and implicit instruction on free written response task performance. *Language Learning*, *61*, 868–903.

Benati, A. (2005). The effects of processing instruction, traditional instruction and meaning–output instruction on the acquisition of the English past simple tense. *Language Teaching Research*, *9*, 67–93.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2005). *Comprehensive meta-analysis* [computer software]. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2011). *Introduction to metaanalysis*. West Sussex: Wiley.

*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.

*Day, E., & Shapson, S. (1991). Integrating formal and functional approaches to language teaching in French immersion: An experimental study. *Language Learning*, *41*, 25–58.

*de Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, *19*, 249–297.

*De Jong, N. (2005). Can second language grammar be learned through listening?: An experimental study. *Studies in Second Language Acquisition*, *27*, 205–234.

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

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

Doughty, C. (2003). Instructed SLA: Constraints, compensation, and enhancement. In Long, M., & C. Doughty (Eds.), *Handbook of second language acquisition* (pp. 257–310). Malden, MA: Blackwell.

Doughty, C., & Williams, J. (1998). Issues and terminology. In Doughty, C., & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 1–12). Cambridge: Cambridge University Press.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*, 141–172.

Ellis, R. (2015). *Understanding Second Language Acquisition: 2nd Edition*. Oxford: Oxford University Press.

*Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*, 339–368.

*Erlam, R. (2003). The effects of deductive and inductive instruction on the acquisition of direct object pronouns in French as a second language. *The Modern Language Journal*, *87*, 242–260.

*Fernández, C. (2008). Reexamining the role of explicit information in processing instruction. *Studies in Second Language Acquisition*, *30*, 277–305.

*Fotos, S., & Ellis, R. (1991). Communicating about grammar: A task-based approach. *TESOL Quarterly*, *25*, 605–628.

Frederickson, N., Reed, P., & Clifford, V. (2005). Evaluating web-supported learning versus lecture-based teaching: Quantitative and qualitative perspectives. *Higher Education*, *50*, 645–664.

*Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, *34*, 445–474.

Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revisited and updated. In Rebuschat, P. (Ed.), *Implicit and explicit learning of languages* (pp. 443–483). Amsterdam: John Benjamins.

*Han, Z.-H. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, *36*, 543–572.

Han, Z-H. (2011). Fossilization – A classic concern of SLA research. In Gass, S., & Mackey, A. (Eds.), *The Routledge handbook of second language acquisition* (pp. 476–490). New York: Routledge.

Han, Z-H. (2014). From julie to wes to alberto: Revisiting the construct of fossilization. In Han, Z-H., & Tarone E. (Eds.), *Interlanguage: Forty years later* (pp. 47–74). Amsterdam: John Benjamins.

Han, Z-H. (2015). Striving for complementarity between narrative and meta-analytic reviews. *Applied Linguistics*, *36*(3), 409–415.

*Harley, B. (1989). Functional grammar in French immersion: A classroom experiment. *Applied Linguistics*, *10*, 331–359.

*Henshaw, F. (2012). How effective are affective activities? Relative benefits of two types of structured input activities as part of a computer-delivered lesson on the Spanish subjunctive. *Language Teaching Research*, *16*, 393–414.

*Hinenoya, K., & Lyster, R. (2015). Identifiability and accessibility in learning definite article usages: A quasi-experimental study with Japanese learners of English. *Language Teaching Research*, *19*, 397–415.

*Iwashita, N. (2003). Negative feedback and positive evidence in task-based interaction. *Studies in Second Language Acquisition*, *25*, 1–36.

*Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis. *Studies in Second Language Acquisition*, *24*, 541–577.

Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal*, *99*, 1–18.

*Kartchava, E., & Ammar, A. (2014). The noticeability and effectiveness of corrective feedback in relation to target type. *Language Teaching Research*, *18*, 428–452.

*Lee, S.-K. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning*, *57*, 87–118.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*, 345–366.

*Leeman, J. (2003). Recasts and second language development. *Studies in Second Language Acquisition*, *25*, 37–63.

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

*Li, Q. (2012). Effects of instruction on adolescent beginners' acquisition of Request Modification. *TESOL Quarterly*, *46*, 30–55.

Li, Shaofeng (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*, 309–365.

*Li, Shaofeng (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, *97*, 634–654.

Li, Shaofeng (2015). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics*, *36*, 385–408.

*Li, Shuai (2012). The effects of input-based practice on pragmatic development of requests in L2 Chinese. *Language Learning*, *62*, 403–438.

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Long, M. (1983). Does second language instruction make a difference?: A review of the Research. *TESOL Quarterly*, *17*, 359–382.

*Long, M.H., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *The Modern Language Journal*, *82*, 357–371.

*Loschky, L. (1994).Comprehensible input and second language acquisition: What is the relationship? *Studies in Second Language Acquisition*, *16*, 303–323.

*Lyddon, P.A. (2011). The efficacy of corrective feedback and textual enhancement in promoting the acquisition of grammatical redundancies. *The Modern Language Journal*, *95*, 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. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, *26*, 399–432.

Mackey, A., & Goo, J.(2007). Interaction research in SLA: A meta-analysis and research synthesis. In Mackey, A. (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 407–452). Oxford: Oxford University Press.

*Marsden, E., & Chen, H.-Y. (2011). The roles of structured input activities in processing instruction and the kinds of knowledge they promote. *Language Learning*, *61*, 1058–1098.

*Morgan-Short, K., & Bowden, H.W. (2006). Processing instruction and meaningful output-based instruction: Effects on second language development. *Studies in Second Language Acquisition*, *28*, 31–65.

*Nagata, N. (1993). Intelligent computer feedback for second language instruction. *The Modern Language Journal*, *77*, 330–339.

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

Plonsky, L. (2016). Applying meta-analysis in instructed L2 research. Unpublished post-conference workshop, The 35th Second Language Research Forum (SLRF 2016), Columbia University, New York. *Abstract* available at: http://www.tc.columbia.edu/slrf2016/workshops (accessed June 2018).

*Rassaei, E. (2014). Scaffolded feedback, recasts, and L2 development: A sociocultural perspective. *The Modern Language Journal*, *98*, 417–431.

*Rassaei, E. (2015). Recasts, field dependence/independence cognitive style, and L2 development. *Language Teaching Research*, *19*, 499–518.

*Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, *31*, 437–470.

*Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*, *62*, 93–132.

*Révész, A., Sachs, R., & Hama, M. (2014). The effects of task complexity and input frequency on the acquisition of the past counterfactual construction through recasts. *Language Learning*, *64*, 615–650.

*Rosa, E., & O'Neill, M. D. (1999). Explicitness, intake and the issue of awareness. *Studies in Second Language Acquisition*, *21*, 511–556.

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In Norris, J.M., & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133–164). Philadelphia, PA: John Benjamins.

*Sagarra, N., & Abbuhl, R. (2013). Optimizing the noticing of recasts via computer-delivered feedback: Evidence that oral input enhancement and working memory help second language learning. *The Modern Language Journal*, *97*, 196–216.

Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition*, *35*(1), 1–29.

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs: Dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research*, *84*(3), 328–364.

*Sheen, Y. (2008). Recasts, language anxiety, modified output, and L2 learning. *Language Learning*, *58*, 835–874.

*Sheen, Y. (2010). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition*, *32*, 203–234.

*Shintani, N. (2012). Input-based tasks and the acquisition of vocabulary and grammar: A process-product study. *Language Teaching Research*, *16*, 253–279.

*Shintani, N., & Ellis, R. (2010). The incidental acquisition of English plural –s by Japanese children in comprehension-based and production-based lessons. *Studies in Second Language Acquisition*, *32*, 607–637.

*Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning*, *64*, 103–131.

Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language Learning*, *63*, 296–329.

Smith, B. (2004). Computer-mediated negotiated interaction and lexical acquisition. *Studies in second language acquisition*, *26*, 365–398.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*, 263–308.

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

*Stefanou, C., & Révész, A. (2015). Direct written corrective feedback, learner differences, and the acquisition of second language article use for generic and specific plural reference. *The Modern Language Journal*, *99*, 263–282.

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, *62*(1), 1–41.

*VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, *15*, 225–241.

*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 effects of positive and negative evidence in the classroom. *Second Language Research*, *7*, 133–161.

*Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese ESL learners' acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition*, *32*, 235–263.

*Yilmaz, Y. (2012). The relative effects of explicit correction and recasts on two target structures via two communication modes. *Language Learning*, *62*, 1134–1169.

Ziegler, N. (2016). Synchronous computer-mediated communication and interaction: A meta-analysis. *Studies in Second Language Acquisition*, *38*, 553–586.

Appendix 1. Results of descriptive coding.

Number	Study (year)	Journal	Age group	Proficiency	SL or FL	Educational setting	Explicit or implicit	Delivery of instruction	Lab vs. classroom
1.	Akakura (2012)	LTR	adult	high	SL	n/a	explicit	computer	classroom
2.	Ammar & Spada (2006)	SSLA	young	n/a	FL	primary	both	face-to-face	classroom
3.	Andringa et al. (2011)	LL	adolescent	low	SL	secondary	both	computer	classroom
4.	Carroll & Swain (1993)	SSLA	adult	mid	SL	language institute	both	face-to-face	lab
5.	Day & Shapson (1991)	LL	adolescent	n/a	SL	secondary	explicit	face-to-face	classroom
6.	de Graaff (1997)	SSLA	adult	low	n/a	university	explicit	computer	lab
7.	De Jong (2005)	SSLA	adult	low	n/a	university	implicit	computer	lab
8.	DeKeyser & Sokalski (1996) (Part 1)	LL	adult	low	FL	university	explicit	pen and paper	classroom
8.	DeKeyser & Sokalski (1996) (Part 2)	LL	adult	low	FL	university	explicit	pen and paper	classroom
9.	Doughty (1991)	SSLA	n/a	mid	SL	language institute	both	computer	lab
10.	Ellis et al. (2006)	SSLA	adult	mid	SL	language institute	both	face-to-face	lab
11.	Erlam (2003)	MLJ	adolescent	mid	FL	secondary	both	face-to-face	classroom
12.	Fernández (2008)	SSLA	adult	mid	FL	university	both	computer	lab
13.	Fotos & Ellis (1991) (Part 1)	TQ	adult	mixed	FL	university	explicit	face-to-face	classroom
13.	Fotos & Ellis (1991) (Part 2)	TQ	adult	mixed	FL	university	explicit	face-to-face	classroom
14.	Goo (2012)	SSLA	adult	mid	FL	university	both	face-to-face	lab
15.	Han (2002)	TQ	n/a	mid	SL	language institute	implicit	face-to-face	classroom
16.	Harley (1989)	AL	young	mid	SL	primary	explicit	face-to-face	classroom
17.	Henshaw (2012)	LTR	adult	mid	FL	university	explicit	computer	lab
18.	Hinenoya & Lyster (2015)	LTR	adult	mixed	FL	graduates	explicit	computer	lab
19.	Iwashita (2003)	SSLA	adult	low	FL	university	both	face-to-face	lab
20.	Izumi (2002)	SSLA	adult	mid	SL	university	implicit	computer	lab
21.	Kartchava & Ammar (2014)	LTR	adult	mid	SL	university	implicit	face-to-face	classroom
22.	Lee (2007)	LL	adolescent	high	FL	secondary	implicit	face-to-face	classroom
23.	Leeman (2003)	SSLA	adult	low	FL	university	implicit	face-to-face	lab
24.	Leow (1998)	MLJ	n/a	low	FL	language institute	both	face-to-face	classroom
25.	Li, Qingping (2012)	TQ	adolescent	low	FL	secondary	both	face-to-face	classroom
26.	Li, Shaofeng (2013)	MLJ	adolescent	mixed	FL	university	both	face-to-face	lab
27.	Li, Shuai (2012)	LL	adult	mid	both	university	implicit	combined	lab
28.	Long et al. (1998) (Part 1)	MLJ	adult	low	FL	university	implicit	face-to-face	lab
28.	Long et al. (1998) (Part 2)	MLJ	adult	mid	FL	university	implicit	face-to-face	lab

(Continued)

Appendix 1. (Continued)

Number	Study (year)	Journal	Age group	Proficiency	SL or FL	Educational setting	Explicit or implicit	Delivery of instruction	Lab vs. classroom
29.	Loschky (1994)	SSLA	adult	mixed	FL	university	implicit	face-to-face	lab
30.	Lyddon (2011)	MLJ	adult	low	FL	university	both	computer	lab
31.	Lyster (1994)	AL	adolescent	high	SL	secondary	both	face-to-face	classroom
32.	Lyster (2004)	SSLA	young	n/a	SL	primary	both	face-to-face	classroom
33.	Marsden & Chen (2011)	LL	young	high	FL	primary	implicit	computer	lab
34.	Morgan-Short & Bowden (2006)	SSLA	mixed	low	FL	university	implicit	computer	lab
35.	Nagata (1993)	MLJ	adult	mid	FL	university	explicit	computer	lab
36.	Rassaei (2014)	MLJ	adult	mid	FL	language institute	both	face-to-face	lab
37.	Rassaei (2015)	LTR	adult	mid	FL	graduate	implicit	face-to-face	classroom
38.	Révész (2009)	SSLA	mixed	mixed	FL	secondary	implicit	combined	lab
39.	Révész (2012)	LL	mixed	low	FL	secondary	implicit	combined	lab
40.	Révész et al. (2014)	LL	adult	mid	SL	university	implicit	face-to-face	lab
41.	Rosa & O'Neill (1999)	SSLA	adult	mid	FL	university	both	pen and paper	lab
42.	Sagarra & Abbuhl (2013)	MLJ	adult	low	FL	university	implicit	computer	lab
43.	Sheen (2008)	LL	adult	mid	SL	university	implicit	face-to-face	classroom
44.	Sheen (2010)	SSLA	n/a	mid	SL	mixed	both	face-to-face	classroom
45.	Shintani (2012)	LTR	young	low	FL	primary	implicit	face-to-face	classroom
46.	Shintani & Ellis (2010)	SSLA	young	low	FL	primary	both	face-to-face	classroom
47.	Shintani et al. (2014)	LL	adult	low	FL	university	explicit	face-to-face	classroom
48.	Stafford et al. (2012)	LL	adult	low	FL	university	explicit	computer	lab
49.	Stefanou & Révész (2015)	MLJ	adolescent	mid	FL	secondary	explicit	face-to-face	classroom
50.	VanPatten & Cadierno (1993)	SSLA	adolescent	mid	FL	secondary	explicit	face-to-face	classroom
51.	VanPatten & Oikkenon (1996)	SSLA	adolescent	mid	FL	secondary	both	face-to-face	classroom
52.	White et al. (1991)	AL	young	low	FL	primary	both	face-to-face	classroom
53.	Yang & Lyster (2010)	SSLA	adult	high	FL	university	both	face-to-face	classroom
54.	Yilmaz (2012)	LL	adult	low	FL	n/a	both	combined	lab

Appendix 1. (Continued)

Number	Study (year)	Total treatment sessions	Length of each treatment session	Intensity of treatment (total treatment)	Duration of instruction (days)	Mode of outcome measures (written vs. oral)	Type of outcome measures
1.	Akakura (2012)	3	less than 60 mins	180 mins	7	combined	MJ, CCR, FCR
2.	Ammar & Spada (2006)	12	30–45 mins	480 mins	28	combined	MJ, CCR, FCR
3.	Andringa et al. (2011)	n/a	60 mins	60 mins	90–120	written	MJ, FCR
4.	Carroll & Swain (1993)	2	n/a	n/a	1	n/a	MJ
5.	Day & Shapson (1991)	n/a	180 mins	n/a	35–49	combined	CCR, FCR
6.	de Graaff (1997)	10	n/a	n/a	35	written	MJ, CCR
7.	De Jong (2005)	4	90 mins	360 mins	14	combined	MJ, SR, CCR
8.	DeKeyser & Sokalski (1996) (Part 1)	5	25–50 mins	250 mins	5	written	SR, CCR
8.	DeKeyser & Sokalski (1996) (Part 2)	5	25–50 mins	250 mins	5	written	SR, CCR
9.	Doughty (1991)	10	n/a	n/a	10	combined	MJ, CCR
10.	Ellis et al. (2006)	2	30 mins	60 mins	2	combined	MJ, CCR
11.	Erlam (2003)	3	45 mins	75 mins	7	combined	CCR, FR
12.	Fernández (2008)	1	n/a	n/a	1	written	SR
13.	Fotos & Ellis (1991) (Part 1)	1	20 mins	20 mins	1	written	MJ
13.	Fotos & Ellis (1991) (Part 2)	1	20 mins	20 mins	1	written	MJ
14.	Goo (2012)	2	20 mins	40 mins	2	written	MJ, CCR
15.	Han (2002)	8	n/a	n/a	60	combined	FCR
16.	Harley (1989)	n/a	n/a	n/a	56	combined	SR, FR
17.	Henshaw (2012)	1	n/a	n/a	1	written	SR
18.	Hinenoya & Lyster (2015)	3	1.5 to 2 hours	6 hours	3	written	CCR
19.	Iwashita (2003)	n/a	n/a	n/a	n/a	oral	FCR
20.	Izumi (2002)	6	n/a	n/a	14	written	MJ, SR, CCR
21.	Kartchava & Ammar (2014)	2	120 mins	240 mins	2	combined	CCR, FCR
22.	Lee (2007)	3	20 mins	60 mins	14	written	CCR
23.	Leeman (2003)	1	20 mins	20	1	oral	FCR
24.	Leow (1998)	1 or 2*	n/a	n/a	1	written	SR, CCR
25.	Li, Qingping (2012)	3	45 mins	135	21	written	CCR
26.	Li, Shaofeng (2013)	1	40–45 mins	45	1	combined	MJ, CCR
27.	Li, Shuai (2012)	2	20–30 mins	50 mins	2	oral	SR, CCR
28.	Long et al. (1998) (Part 1)	1	n/a	40 mins	1	oral	CCR

(Continued)

Appendix 1. (Continued)

Number	Study (year)	Total treatment sessions	Length of each treatment session	Intensity of treatment (total treatment)	Duration of instruction (days)	Mode of outcome measures (written vs. oral)	Type of outcome measures
28.	Long et al. (1998) (Part 2)	1	n/a	40 mins	1	combined	CCR
29.	Loschky (1994)	3	15–30 mins	90 mins	3	written	SR
30.	Lyddon (2011)	1	n/a	n/a	1	written	CCR
31.	Lyster (1994)	n/a	n/a	n/a	n/a	combined	SR, FCR
32.	Lyster (2004)	n/a	9 hours	n/a	35	combined	SR, CCR, FCR
33.	Marsden & Chen (2011)	4	40 mins	160 mins	4	combined	MJ, CCR, FCR
34.	Morgan-Short & Bowden (2006)	1	1 hour	60 mins	1	written and aural	SR, CCR
35.	Nagata (1993)	6	n/a	n/a	7	written	CCR
36.	Rassaei (2014)	3	14 to 24 mins	60 mins	7	combined	MJ, FCR
37.	Rassaei (2015)	4	35 mins	135 mins	7	combined	FCR
38.	Révész (2009)	3	15 mins	45	1	combined	FCR
39.	Révész (2012)	3	15 mins	45	1	combined	MJ, FCR
40.	Révész et al. (2014)	2	20–25 mins	50	14	written	SR, FCR
41.	Rosa & O'Neill (1999)	1	n/a	n/a	1	written	SR
42.	Sagarra & Abbuhl (2013)	1	50 mins	50	35	combined	SR, CCR, FCR
43.	Sheen (2008)	2	n/a	n/a	21	written	MJ, CCR, FCR
44.	Sheen (2010)	2	5 to 30 mins	60	1	written	MJ, CCR, FCR
45.	Shintani (2012)	9	n/a	n/a	35	combined	SR, CC
46.	Shintani & Ellis (2010)	6	30–40 mins	240	21	combined	SR, CC
47.	Shintani et al. (2014)	1	5 to 20 mins	20 mins	1	written	FCR
48.	Stafford et al. (2012)	1	n/a	n/a	1	written and aural	MJ, SR, CCR
49.	Stefanou & Révész (2015)	2	10 mins	20	2	written	SR, FCR
50.	VanPatten & Cadierno (1993)	2	n/a	n/a	2	written	SR, CCR
51.	VanPatten & Oikkenon (1996)	4	n/a	n/a	4	written and aural	SR, CCR
52.	White et al. (1991)	n/a	8 hours	n/a	14	combined	MJ, CCR, FCR
53.	Yang & Lyster (2010)	n/a	120 mins	n/a	14	combined	FR
54.	Yilmaz (2012)	2	n/a	n/a	n/a	combined	SR, CCR

Notes. *1 for a single exposure group, 2 for a multiple exposure group. AL = *Applied Linguistics*; LL = *Language Learning*; LTR = *Language Teaching Research*; MLJ = *The Modern Language Journal*; SSLA = *Studies in Second Language Acquisition*; TQ = *TESOL Quarterly*. FL = foreign language; SL = second language. CCR = constrained constructed response; FCR = free constructed response; MJ = metalinguistic judgment; SR = selected response. FL = foreign language; SL = second language

Appendix 2. A summary of the effect sizes of the individual studies.

Primary studies	Effect size (*g*)	Standard error	95% CI lower limit	95% CI upper limit
1. Akakura (2012)	0.64	0.21	0.22	1.05
2. Ammar & Spada (2006)	0.65	0.43	–0.19	1.48
3. Andringa de Glopper & Hacquebord (2011)	0.43	0.25	–0.06	0.98
4. Carroll & Swain (1993)	1.51	0.36	0.81	2.21
5. Day & Shapson (1991)	0.92	0.12	0.69	1.16
6. de Graff (1997)	1.57	0.22	1.14	1.98
7. De Jong (2005)	–0.22	0.32	–0.86	0.41
8. DeKeyser & Sokalski (1996)	3.77	0.68	2.44	5.09
9. Doughty (1991)	0.19	0.52	–0.84	1.21
10. Ellis Loewen & Erlam (2006)	0.35	0.42	–0.48	1.17
11. Erlam (2003)	0.81	0.31	0.21	1.42
12. Fernández (2008)	0.41	0.29	–0.16	0.98
13. Fotos & Ellis (1991)	1.98	0.45	1.09	2.86
14. Goo (2012)	1.22	0.44	0.36	2.08
15. Harley (1989)	3.55	1.04	1.51	5.56
16. Han (2002)	2.57	0.90	0.82	4.33
17. Henshaw (2012)	0.34	0.26	–0.16	0.84
18. Hinenoya & Lyster (2015)	0.67	0.22	0.23	1.10
19. Iwashita (2003)	0.77	0.31	0.15	1.86
20. Izumi (2002)	0.52	0.29	–0.04	1.09
21. Kartchava & Ammar (2014)	0.29	0.30	–0.31	0.88
22. Lee (2007)	1.05	0.18	0.69	1.41
23. Leeman (2003)	1.52	0.40	0.74	2.30
24. Leow (1998)	–0.12	0.30	–0.70	0.46
25. Li, Qingping (2012)	–0.01	0.18	–0.36	0.33
26. Li, Shaofang (2013)	1.0	0.25	0.50	1.50
27. Li, Shuai (2012)	1.2	0.47	0.29	2.13
28. Long, Inagaki & Ortega (1998)	1.05	0.52	0.02	2.07
29. Loschky (1994)	0.18	0.38	–0.56	0.91
30. Lyddon (2011)	2.87	0.95	2.19	3.56
31. Lyster (1994)	1.21	0.22	0.77	1.64
32. Lyster, R. (2004)	1.24	0.32	0.62	1.87
33. Marsden, E., & Chen, H.-Y. (2011)	0.89	0.38	0.15	1.62

Appendix 2. (Continued)

Primary studies	Effect size (*g*)	Standard error	95% CI lower limit	95% CI upper limit
34. Morgan-Short & Bowden (2006)	1.78	0.45	0.90	2.65
35. Nagata (1993)	0.27	0.34	–0.39	0.93
36. Rassaei, E. (2014)	1.82	0.34	1.16	2.48
37. Rassaei, E. (2015)	0.77	0.33	0.14	1.41
38. Révész, A. (2009)	6.19	0.80	4.60	7.72
39. Révész, A. (2012)	3.95	0.50	2.96	4.93
40. Révész, Sachs, & Hama (2014)	0.69	0.44	–0.17	1.55
41. Rosa & O'Neill (1999)	–0.19	0.38	–0.94	0.57
42. Sagarra & Abbuhl (2013)	3.89	0.47	2.97	4.81
43. Sheen (2008)	0.19	0.43	–0.67	1.04
44. Sheen (2010)	0.47	0.27	–0.06	0.99
45. Shintani (2012)	0.98	0.39	0.22	1.74
46. Shintani & Ellis (2010)	2.05	0.50	1.07	3.03
47. Shintani, Ellis, Suzuki (2014)	1.37	0.30	0.77	1.96
49. Stafford, Bowden & Sanz (2012)	1.08	0.40	0.29	1.88
49. Stefanou & Révész (2015)	0.46	0.27	–0.06	0.98
50. VanPatten & Cadierno (1993)	1.75	0.32	1.12	2.38
51. VanPatten & Oikkenon (1996)	–0.70	0.33	–1.14	–0.06
52. White et al. (1991)	1.15	0.23	0.70	1.60
53. Yang & Lyster (2010)	0.47	0.30	–0.12	1.06
54. Yilmaz (2012)	0.98	0.37	0.25	1.70