SSLA, **30**, 307–331. Printed in the United States of America.
doi:10.1017/S0272263108080479

VISUAL INPUT ENHANCEMENT AND GRAMMAR LEARNING

A Meta-Analytic Review

Sang-Ki Lee and Hung-Tzu Huang
University of Hawai'i at Mānoa

Effects of pedagogical interventions with visual input enhancement on grammar learning have been investigated by a number of researchers during the past decade and a half. The present review delineates this research domain via a systematic synthesis of 16 primary studies (comprising 20 unique study samples) retrieved through an exhaustive literature search. The overall magnitude of visual input enhancement was addressed by calculating and aggregating effect size *d* values. The results indicate that second language readers provided with enhancement-embedded texts barely outperformed those who were exposed to unenhanced texts with the same target forms flooded in them ($d = 0.22$). A theoretical tension between form and meaning was indicated by a small but negative effect size value for learners' meaning processing ($d = -0.26$). The importance of improving methodological practices in this research domain, including the reporting of statistical and treatment-related information and the counteracting of a possible publication bias, was also revealed by the synthetic analyses and is further discussed.

Ever since a large consensus was reached on the benefits of second language (L2) instruction in the field of instructed SLA (Doughty, 1991; Long, 1991; Norris

A version of this article was presented at the American Association for Applied Linguistics Annual Conference, Costa Mesa, California, April 21, 2007. We would like to thank Lourdes Ortega for her invaluable help throughout the research project. We are also grateful to four anonymous *SSLA* reviewers for their helpful comments.

Address correspondence to: Sang-Ki Lee or Hung-Tzu Huang, Department of Second Language Studies, University of Hawai'i at Mānoa, 1890 East-West Road, Honolulu, HI 96822; e-mail: sangki@hawaii.edu or hungtsu@hawaii.edu.

© 2008 Cambridge University Press 0272-2631/08 $15.00

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

& Ortega, 2000), much effort has been devoted to addressing the relative effectiveness of various pedagogical interventions. In the same vein, a growing body of visual input enhancement (VIE; also known as written or textual enhancement) studies has tapped the possibility of making input more perceptible to L2 learners by employing enhancement techniques with typographical cues (e.g., *underlining*, **boldfacing**, *italicization*, CAPITALIZATION, or other strategies such as color coding or using different font sizes or types). In this way, enhancement-embedded texts are used to make a target form perceptually salient to the extent that processing that form is facilitated. Schmidt's (2001) noticing hypothesis offers rationales for such claims: For input to be processed for acquisition by L2 learners, it must first be noticed. Enhancing input using typographical techniques increases the chance that the visually prominent input will be noticed and will thus establish a trace in long-term memory.

Among studies investigating VIE, some have treated it as the main independent variable of interest (e.g., Jourdenais, 1998; Leow, 2001) and some have used it as one of several techniques for focus on form (e.g., Doughty, 1991; Robinson, 1997). Others have included an examination of additional moderating variables, such as explicit directions to focus on form (Shook, 1994, 1999), explicit rule instruction prior to exposure to the enhanced input (Alanen, 1995; Kubota, 2000), communicative tasks (Leeman, Arteagoitia, Fridman, & Doughty, 1995), text length (Leow, 1997), extensive reading and listening (White, 1998), output activities (Izumi, 2002), relative communicative value of form (Overstreet, 2002), relative saliency of form (Leow, Egi, Nuevo, & Tsai, 2003), simplified modification (Wong, 2003), opportunity of online chatting (Ha, 2005), and topic familiarity (Lee, 2007; Overstreet, 1998).

To date, accumulated findings and conclusions drawn by individual researchers are elusive. All previous studies, with the single exception of Jourdenais's (1998) doctoral research, utilized comparison groups rather than true control groups: By and large, VIE has been compared to the more implicit technique of input flood, or the provision of as many instances of the targeted L2 form in the input as possible. Some of the studies found comparatively positive effects of VIE over input flood on learning of grammatical items (e.g., Lee, 2007; White, 1998), whereas others failed to find such helpful effects (e.g., Izumi, 2002, 2003; Jourdenais). Explicit grammar instruction, to which VIE has also frequently been compared, has typically been reported to have superior effects to implicit VIE instruction (e.g., Alanen, 1995; Doughty, 1991). Some researchers have also shown disagreement on the effects of VIE on learners' noticing. Based on two empirical studies, Izumi (2002, 2003) maintained that VIE was beneficial for promoting learners' noticing, even though he did not find any benefit of such noticing in terms of actual learning gains. Conversely, two other studies (Leow, 2001; Leow et al., 2003) concluded that VIE does not aid learners' enhanced degree of noticing.

Existing primary studies have shown a large divergence in terms of their study features. For example, some researchers have studied the effects of VIE not only on acquisition but also on comprehension (Doughty, 1991; Jour-

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

denais, 1998; Lee, 2007; Leow, 1997, 2001; Leow et al., 2003; Overstreet, 1998, 2002; Wong, 2003). Although most of these studies did not report statistically significant hampering effects of VIE on comprehension, both Lee and Overstreet (1998) reported that VIE negatively affected learners' meaning comprehension. In the context of a psycholinguistic discussion about learner attention and the relation of form and meaning, Lee argued that learners' comprehension might be negatively affected by attention-drawing VIE techniques and proposed that a balanced amount of intervention to teach L2 forms during reading would be needed. Investigating the potential effects of VIE on learners' meaning processing is indeed a necessary part of the VIE research program in order to adequately understand the role of VIE in grammar pedagogy.

Considering the diverse study features and conflicting results from previous studies, we feel that the accumulating research in this domain warrants a comprehensive synthesis of study findings. By simply following the vote-counting manner of review (Light & Pillemer, 1984), one might tally and compare the number of studies demonstrating beneficial and nonbeneficial results to come to a conclusion about the effectiveness of VIE as a pedagogical choice for grammar instruction. Many research synthetists, however, caution that this traditional practice might obscure the impact of study features across individual primary studies and the impact of the magnitude of each study.

Employing the systematic procedures offered by the methodology of quantitative meta-analytic synthesis (Norris & Ortega, 2006), the current meta-analysis brings together previous studies to examine VIE research and determine whether VIE has beneficial effects for grammar learning. By synthesizing the available empirical research in the field, this methodology not only enables us to examine the overall effectiveness of VIE but also allows us to investigate other important research questions not addressed in the primary research. For example, many meta-analysts admonish the prevailing publication practices that favor studies reporting statistically significant findings (Cooper, 1998; Rosenthal, 1994; see also Norris & Ortega, 2000, in particular, for a discussion on publication bias in the SLA field). Thus, it might be important to ask what the relative effects of VIE are, depending on the sources of publication as a moderating variable in this secondary research. Light and Pillemer (1984) stressed that research synthesis allows us to "learn from existing findings, to discover what is known" (p. ix). A meta-analysis approach, then, attempts to shed new light on available data in primary studies via systematic interpretations. Four research questions guided the present study:

1. How has VIE been investigated in the SLA literature to date?
2. What is the magnitude of the impact of VIE on learners' grammar learning?
3. What is the magnitude of the impact of VIE on learners' meaning comprehension?
4. What is the relative magnitude of the impact of VIE in published versus unpublished studies? Is there a difference in relative magnitude observable between studies published in refereed journals and unpublished doctoral studies?

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

The first research question was addressed via a systematic synthesis of study features and the other three research questions were addressed through the lens of a quantitative meta-analysis.

METHOD

The Literature Search

An extensive and principled search of the literature was performed for the current meta-analysis by exploring online databases and checking references as well as footnotes of relevant journals and review articles. Our aim at this stage was to trace all accessible SLA studies that empirically examined the effects of VIE on learning of grammatical items through reading tasks.

To our knowledge, Lee (2007) is the most recent study that attempted to undertake a comprehensive review of the VIE research domain. Of the 13 studies reviewed there, 11 were published articles and 2 were doctoral dissertations. With this work as a starting point, we undertook a further search for published articles and unpublished dissertations on the effects of VIE on grammar learning in five electronic databases: Educational Resources Information Center, Linguistic and Language Behavior Abstracts, ProQuest Digital Dissertations, PsycARTICLES, and PsychInfo. Combinations of the following terms were used in our search for abstracts in the databases: (a) OR search terms: input enhancement, enhanc*, typographical, textual, focus on form, implicit, unobtrusive, incidental, attention, attend, and notic*; (b) AND search terms: reading, grammar, and linguistic. In our attempt to conduct a comprehensive and exhaustive search of the available resources, we also manually checked five journals that have published at least one VIE study (*Applied Language Learning*, *Hispania*, *Language Learning*, *Spanish Applied Linguistics*, and *Studies in Second Language Acquisition*), from the first issue of 1981 to the latest issue as of April 2006.[1] We then extended the manual search to four additional journals (*Applied Linguistics*, *Modern Language Journal*, *Second Language Research*, and *TESOL Quarterly*) as well as references and footnotes in relevant textbooks (e.g., Doughty & Williams, 1998; Wong, 2005). Through these extensive and systematic searches, two additional studies were identified as eligible for the current research synthesis (Ha, 2005; Kubota, 2000). As a result, a total of 16 primary studies were identified and analyzed for the present synthesis. All the studies included are marked with an asterisk (*) in the Reference section. Upon coding and analysis of each primary study, it was determined that the 16 studies contributed a total of 20 unique samples available for synthesis; that is, when different groups of learners (i.e., different unique samples) within the same study received VIE treatments on distinct target forms, we treated each group as a unique study sample and extracted an effect size for each one (e.g., Jourdenais, 1998; Kubota; Overstreet, 2002).

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

It should be noted here that we included only published studies and dissertations. Thus, it is possible that there might be a large pool of studies that have not been published and are excluded from the current synthesis. For example, in the process of the literature search we encountered some so-called fugitive literature, such as manuscripts in conference proceedings (e.g., Wijaya, 2000) and in-house working papers (e.g., Izumi, 2003). We decided not to include fugitive literature other than dissertations because of considerations of replicability of the current research synthesis, as it was evidently infeasible to locate all such literature with the methods that we undertook for study retrieval. We would like to emphasize that in no way is the present study a final, definitive answer to the questions regarding the effects of VIE. Rather, the study should be understood as a formative evaluation of this research domain.

Study Eligibility Criteria

The primary studies retrieved for the current meta-analysis met the following inclusion and exclusion criteria:

1. Studies included in the current meta-analysis were experimental or quasi-experimental in design, recruiting participants who were L2 or foreign language learners. First language (L1) studies were excluded from the current review (e.g., Lorch, Lorch, & Klusewitz, 1995).
2. Studies included examined the effects of VIE on learning of grammatical items targeted by study design through reading tasks, and the effects were investigated by posttest measure(s). Studies involving single-sentence reading tasks rather than natural reading tasks (e.g., Robinson, 1997; J. N. Williams, 1999) and those that examined the development of vocabulary knowledge (e.g., Barcroft, 2003; Kim, 2003) or L2 pragmatic awareness (e.g., Witten, 2002) were therefore excluded.
3. Studies included compared the relative effects of VIE with control or comparison group(s). Studies in which VIE groups served as comparison groups to address questions other than the effects of VIE were therefore excluded (e.g., Sato, 2005; Seiba, 2001).
4. Studies included were articles published in refereed journals, articles published as book chapters, or unpublished dissertations. Other fugitive literature, such as in-house working papers (e.g., Izumi, 2003), papers in conference proceedings (e.g., Wijaya, 2000), and master's theses (e.g., Pacheco, 2004), was excluded.
5. Studies included reported adequate information for effect sizes to be calculated. Thus, studies that failed to report descriptive statistics were not included (e.g., Leeman et al., 1995).
6. Studies included appeared between the years 1981 and 2006 (see Note 1).
7. Studies included were written in English. This was to avoid any possible bias in sampling the pool of primary studies and thereby to ensure the replicability of the current meta-analysis.
8. If the same study and sample was reported across several sources, only one report was included in the current meta-analysis. For example, if a dissertation and a published article based on it existed, the latter was included in the synthesis (i.e.,

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

Doughty, 1988, 1991; Izumi, 2000, 2002; White, 1996, 1998; Wong, 2000, 2003), and the dissertation was consulted to retrieve data necessary for the meta-analysis only if these data were not reported in the published article.

Coding of Primary Studies

After identifying the body of research literature that met our inclusion criteria, a coding scheme was developed to classify common characteristics of the 16 studies and to compile the relevant information from all of them. After extensive piloting, the final comprehensive coding scheme included two major categories for methodological features: (a) learner characteristics and (b) research design. The complete list of variables considered for coding is included in the Appendix.

Learner characteristics were considered a main category for the analysis, so studies were coded for total number of participants, number of participants per each unique study sample, length of study of the target language, and context of L2 study. Information on the extent of previous knowledge of the participants on the targeted form and the actual proficiency levels of participants were also coded, because these attributes were considered critical factors in determining the impact of VIE. Primary studies described the proficiency levels of participants in various ways. To aid in coding proficiency, we employed Thomas's (2006) four-way classification for proficiency assessment measures: (a) impressionistic judgment, (b) institutional status, (c) in-house assessment, and (d) standardized tests (see Thomas for details of each assessment type).

As for the design features of the primary studies, we coded such features as the target language, targeted linguistic forms, reading materials, dependent variables and types of measures used, and any other independent variables examined apart from VIE. With regard to the features of dependent variables in particular, we coded the relative processing demands the measures placed on the participants as they engaged in producing language. The measures were first distinguished as receptive or productive. For instance, one of the most common measures used for pretests, the grammaticality judgment task, is typical of a receptive measure. Picture description, often used for posttests, was coded as a productive measure. If the measure was considered to tap the productive side of language use, the coders then determined how much guidance was provided for such production. A free written recall, for example, was considered less guided and was thus classified as a free constructed measure, whereas a short-answer type of comprehension task was regarded as more guided and was therefore coded as a constrained constructed measure. We also coded the measures used for gauging learners' degree of noticing.

Additionally, we reasoned that the intensity of treatment sessions—namely the amount of exposure to target forms—would potentially account for the

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

outcomes of research findings. Not only the number of treatment sessions but also the detailed amount of time for each step of the procedure in the experiments were therefore coded, including the interval between the pretest and the first treatment session, the interval between the last treatment session and the posttest, the duration of the treatments, the amount of time for all treatment sessions, and the actual reading time.

The two authors worked independently to code each study according to the established criteria. The intercoder reliability estimate between the two researchers was 94.1% simple percent agreement. Any disagreements were resolved through iterative discussion until 100% agreement was obtained.

RESULTS

The Research Synthesis

We examined the 12 published articles and 4 unpublished dissertations according to the established coding categories to obtain a comprehensive understanding of the characteristics of the studies.

Publication Characteristics. Although Sharwood Smith (1981) first suggested the potential pedagogical benefits of VIE, it was not until 1991—10 years later—that the first empirical research designed to probe its actual effects appeared (Doughty, 1991). During our literature search, we noted that since the advent of such VIE studies, only a small number of studies have been conducted (16 since 1991). Furthermore, only a small number of researchers have been engaged in investigating the use of VIE as a pedagogic choice (a total of 16), and we found some researchers as authors of more than two of the studies we located for inclusion (e.g., Doughty and Leow for three studies each). This limited number of studies from a small research community resulted in a great deal of the overlap between the dissertations and published articles that we found: As shown in Figure 1, of the eight dissertation studies that met our inclusion and exclusion criteria after the screening procedure, four were already included in the published articles eligible for inclusion in the meta-analysis. This publication pattern indicates that the VIE domain is still a young area of research.

Learner Characteristics. A total of 1,257 participants and 20 unique samples were involved in the 16 VIE studies, most of which were from foreign language contexts ($n = 13$, or 81% of the 16 studies). As can be seen from the descriptive statistics of the number of participants across each study in Table 1, a wide range of n sizes across the studies was obvious, with a minimum of 10 and a maximum of 256 participants within a single study. The mean and median numbers of total participants were 62.85 and 48.50, respectively. Experimental groups tended to have a larger number of participants than comparison groups.

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

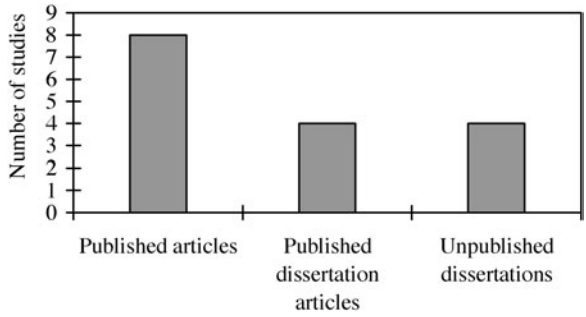


Figure 1. Sources of publication for the included studies.

Thomas's (2006) categorization of learner proficiency measures was adopted to code how researchers documented the proficiency levels of participants. It was found that studies in the VIE domain exclusively consulted the institutional status of their participants for estimating proficiency levels. We also examined actual learner proficiency and reported years of L2 instruction. As illustrated in Figure 2, only one study (Jourdenais, 1998) involved advanced learners (6%), whereas four and nine studies recruited beginning and intermediate learners (25% and 56%), respectively. The remaining two studies did not provide sufficient information on either the length of language instruction or the institutional status of their participants (13%). This pattern led us to presume that primary researchers regarded learners at intermediate levels appropriate for their studies. Beginners might not have acquired sufficient reading skills required for processing VIE-embedded texts and, conversely, advanced learners might have had abundant previous exposure to various

Table 1. Sample size across study and groups

Measure	Full study (n = 16)	Experimental group (k = 20)	Comparison group (k = 16)
Mean	62.85	31.20	26.65
SD	53.56	27.60	25.97
Median	48.50	25.50	21.50
Maximum	256	132	127
Minimum	10	5	5
Range	246	127	122
Total	1,257	624	533

Note. k = number of samples contributing to the effect size calculation.

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

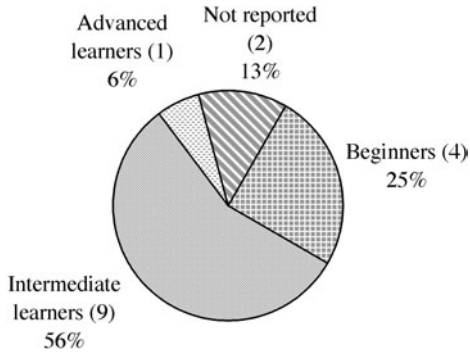


Figure 2. Proficiency levels of study participants.

grammatical structures, which might result in an intervening variable influencing the findings on VIE effectiveness.

It was expected that learners' prior knowledge of target forms would significantly affect the degree of acquisition of enhanced grammatical forms. Based on reported pretest scores, learners from 13 of the 16 studies (81%) were determined to have little or minimal knowledge of the target forms under investigation. Of the three remaining studies (19%), one (Alanen, 1995) targeted artificial Finnish and thus presumed no prior knowledge of the language on the part of the participants. In the other two studies (Ha, 2005; Jourdenais, Ota, Stauffer, Boyson, & Doughty, 1995), participants scored more than 60% on the pretests. Ha's study targeted English determiners, linguistic elements with presumably high frequency, which might explain participants' relatively high degree of previous knowledge of the target forms. Learners in Jourdenais et al.'s study scored an average of 81% on pretests of Spanish preterit and imperfect tense. One should employ caution, however, when interpreting a study with learners who already possess a rich knowledge of the target prior to the study's treatments.

Research Design. Eight of the 16 studies had Spanish (50%) and 5 had English (31%) as their L2 target; the remaining 3 studies targeted Finnish, French, and Japanese, respectively. As for the target forms examined, the 16 primary studies employed a wide range of grammatical structures as the focus of enhancement (see Table 2). It seems that decisions about which structures to target mainly depended on the researchers' judgments about which grammatical structures cause difficulty for learners. Additionally, half of the studies focused on only one target form (n = 8) and the other half utilized two different target forms concurrently highlighted in the texts for learners to process (n = 8). Among the eight studies designed with two target forms, five studies presented the two target forms to the same group of participants and three studies recruited different VIE groups for different target forms.

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

Table 2. Targets of enhancement across the 16 studies

Study by number of target forms	Target language	Enhanced L2 target
One target form		
Doughty (1991)	English	Relativization
Ha (2005)	English	Determiners
Izumi (2002)	English	Relative clauses
Lee (2007)	English	Passive voice
Leow (1997)	Spanish	Impersonal imperative
Leow (2001)	Spanish	Formal/polite imperative
White (1998)	English	Third person singular possessive determiners
Wong (2003)	French	Past participle agreement in relative clauses
Two target forms		
Alanen (1995)	Artificial Finnish	(a) Locative suffix, (b) consonant alternation
Jourdenais (1998)	Spanish	(a) Preterit, (b) imperfect
Jourdenais et al. (1995)	Spanish	(a) Preterit, (b) imperfect
Kubota (2000)	Japanese	(a) Gerund adjectives, (b) gerund of verbs
Leow et al. (2003)	Spanish	(a) Present perfect, (b) present subjunctive
Overstreet (1998)	Spanish	(a) Preterit, (b) imperfect
Overstreet (2002)	Spanish	(a) Gerund adjective, (b) gerund of verbs
Shook (1994)	Spanish	(a) Present perfect, (b) relative pronouns

The most commonly employed VIE cues were boldfacing and underlining (n = 14 and 11, respectively). One study employed an italicization method as the only typographical cue (Alanen, 1995), and the others utilized more than two cues in combination and up to four cues concurrently.

Half of the studies utilized only one measure to determine the extent of learners' grammar learning (n = 8) and the other half used two or more measures (n = 8). As shown in Figure 3, over half (59%) of a total of 22 grammar learning measures were designed to elicit productive responses (n = 13) rather than to test learners' receptive knowledge (n = 9, or 41%). Nine of the 13 productive measures, such as sentence completion tasks, fill-in-the-blank production tasks, and correction tasks, were considered constrained construction types (69%) and four productive measures were free construction types (31%) such as written narration tasks and oral picture description tasks.

Nine studies (Doughty, 1991; Jourdenais, 1998; Lee, 2007; Leow, 1997, 2001; Leow et al., 2003; Overstreet, 1998, 2002; Wong, 2003) further assessed learners' degree of comprehension while reading (56%), mainly employing either

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

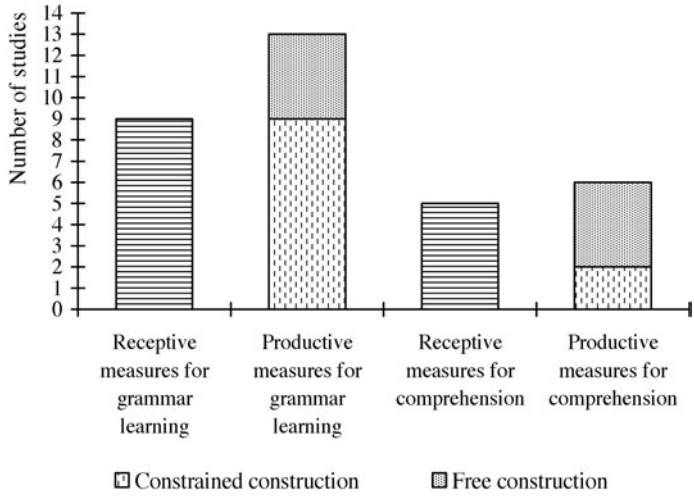


Figure 3. Measures of grammar learning and meaning comprehension.

various forms of comprehension questions or free recalls as measures. Figure 3 indicates that of the 11 reading comprehension measures used across the 9 studies, 5 (45%) were receptive measures such as multiple-choice comprehension questions, and the remaining 6 (55%) were productive measures: 4 free constructions (all of them L1 free recall tasks) and 2 constrained constructions in short-answer comprehension checkup formats. Although the reported effects of VIE on learners' meaning processing were not statistically significant in most of the studies, both Lee and Overstreet (1998) reported that VIE might have observable debilitating effects on the processing of meaning comprehension.

In addition to grammar learning and reading comprehension, the amount of noticing of target forms induced by VIE, as a moderating variable, was another focus of interest in the VIE domain. Among the 16 studies, 3 measured the amount of learners' noticing of the enhanced targets with an explicitly stated research question. Izumi's (2002) study used note-taking scores to investigate whether input enhancement and output activities, in isolation or in combination, can promote noticing. Izumi found that VIE promoted noticing; however, VIE learners who took more notes did not necessarily perform better on L2 posttests. Both Leow (2001) and Leow et al. (2003) employed think-aloud protocols to investigate the benefits of VIE interventions on the degree of noticing. Their results uncovered no helpful effects of VIE on readers' intake or any significant benefits of VIE over unenhanced input for the amount of reported noticing. Although the relative magnitude of VIE effects for grammar pedagogy contingent on noticing measures as well as various target forms

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

would be of interest for further meta-analytic inquiry, these were not amenable to the current meta-analysis because too few studies have looked at noticing and there is no systematic replication in the L2 targets across studies to date.

Treatment Intensity. Overall, the treatments of all 16 primary studies lasted from 1 day to 4 weeks. The number of treatment sessions differed greatly, from 1 single session up to 10 sessions. More specifically, as demonstrated in Table 3, six studies were one-shot experimental designs (38%), six studies administered two or three treatment sessions (38%), and the remaining four studies had more than six treatment sessions (25%).

As seen in Table 3, the difference in the number of treatment sessions directly influenced such factors as the length of the reading texts used, duration of the treatments, total treatment time, and actual exposure time to targeted forms. Thus, the four studies that had one single treatment session lasted less than 30 min (31%). The five studies that had two or three sessions lasted 30 min to less than 2 h (38%), and the other four studies that had more than six sessions are the ones characterized by treatment lengths of more than 4 h (31%). It should be noted that in three studies it was not possible to estimate the actual treatment time (cf. all cases marked n.r. in Table 3) due to insufficient description of the treatment settings. This suggests the need for more explicit, accurate description of study features for future studies. Depending on the number of treatment sessions in each study, there was also a notable difference in the amount of reading in which participants engaged. The longest reading text contained a total of 6,640 words (Ha, 2005) and the shortest was 185 words (Alanen, 1995). Influenced by the nature of target forms and the length of reading texts, the number of tokens per target form also varied substantially, with a maximum of 1,070 (Ha) and a minimum of six exposure instances (Shook, 1994).

Of the 11 studies that reported the interval between the pretest and the first treatment session, 9 studies (82%) had an interval of more than 5 days, presumably to avoid priming effects, and the remaining 2 studies (18%) left less than a 1–2-day interval. Conversely, 11 of the 16 studies (69%) delivered the posttest immediately after the last treatment session. Only three studies (19%) investigated learners' retention of the target form by using delayed posttests, as few as 2 weeks and as many as 5 weeks after the immediate posttest had been administered.

The Quantitative Meta-Analysis

To address the second research question and examine the magnitude of effects of VIE on grammar learning, Cohen's (1988) effect size *d* values were calculated from each individual study. These are presented in Table 4.[2] The main effect size calculation was based on the contrast of the mean of the experimental group with that of the comparison group on immediate posttest mea-

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

Table 3. Treatment sessions and total amount of treatment time

Study by number of sessions	Treatment duration	Reading length	Total exposure
One session			
Jourdenais et al. (1995)	Less than 30 min	205 words	Two forms: 18 and 10 tokens
Leow (1997)	n.r.	Long text: 631 words Short text: 384 words	One form: 15 tokens
Leow (2001)	Less than 30 min	242 words	One form: 15 tokens
Leow et al. (2003)	n.r.	n.r.	One form: 10 tokens
Overstreet (1998)	Less than 30 min	210 words	Two forms: 18 and 10 tokens
Overstreet (2002)	Less than 30 min	588 words	Two forms: 10 tokens each
Two to three sessions			
Alanen (1995)	n.r.	185 words	Two forms: 38 and 23 tokens
Jourdenais (1998)	30 min to less than 2 h	1,589 words	Two forms: 72 tokens each
Kubota (2000)	30 min to less than 2 h	248 words	Two forms: 48 tokens each
Lee (2007)	30 min to less than 2 h	More than 1,200 words	One form: 36 tokens
Shook (1994)	30 min to less than 2 h	402 words	Two forms: 6 tokens each
Wong (2003)	30 min to less than 2 h	1,508 words	One form: 48 tokens
More than six sessions			
Doughty (1991)	More than 4 h	n.r.	n.r.
Ha (2005)	More than 4 h	6,640 words	One form: 1,070 tokens
Izumi (2002)	More than 4 h	840 words	One form: 46 tokens
White (1998)	More than 4 h	n.r.	n.r.

Note. n.r. = not reported

Table 4. Summary of effect size *d* values of primary studies

	Effect size (*d*)			
	Grammar Learning			
Study	Immediate posttest	Delayed posttest	Pre-to-post contrast	Meaning comprehension
Alanen (1995)	0.08	n.t.	n.t.	n.t.
Doughty (1991)	0.46	n.t.	1.85	n.a.
Ha (2005)	0.07	n.t.	0.26	n.t.
Izumi (2002)	0.02	n.t.	0.67	n.t.
Jourdenais (1998) 1	−0.04	−0.10	−0.04	n.a.
Jourdenais (1998) 2	−0.02	−0.10	−0.19	n.a.
Jourdenais (1998) 3	−0.15	0.09	0.01	n.a.
Jourdenais et al. (1995)	1.61	n.t.	n.t.	n.t.
Kubota (2000) 1	−0.37	−0.45	n.a.[a]	n.t.
Kubota (2000) 2	−0.10	−0.21	n.a.	n.t.
Lee (2007)	1.05	n.t.	1.11	−0.73
Leow (1997)	−0.06	n.t.	0.86	0.28
Leow (2001)	n.a.	n.t.	n.a.	n.a.
Leow et al. (2003)	n.a.	n.t.	n.a.	n.a.
Overstreet (1998)	0.07	n.t.	0.07	−0.94
Overstreet (2002) 1	0.61	n.t.	n.t.	−0.34
Overstreet (2002) 2	−0.05	n.t.	n.t.	−0.13
Shook (1994)	n.a.	n.t.	n.a.	−0.30[b]
White (1998)[c]	0.26	−0.01	0.97	n.t.
Wong (2003)	0.23	n.t.	0.45	0.36
Average effect size	0.22	−0.13	0.55	−0.26
95% CI upper	0.47	0.07	0.97	0.19
95% CI lower	−0.03	−0.33	0.13	−0.70
k	17	6	11	7
SD	0.49	0.19	0.62	0.48

Note. n.t. = not tested in the particular study; n.a. = not applicable because the study did not provide sufficient data for the effect size calculation; *k* = number of samples contributing to the effect size calculation.
[a]We could not calculate pre-to-post *d* values for Kubota's (2000) study because different measures were used for the pretest and posttest.
[b]Shook (1999) contributed to the calculation of this value.
[c]The *d* values were calculated from White's (1996) dissertation.

sures. To answer the question adequately, however, the comparisons should have involved true control groups with reading conditions in which the form in focus occurred infrequently in intact texts. This was not possible because there was only one study that met this design criterion (Jourdenais, 1998). Additionally, although means and standard deviations should be available for deriving *d* values, this was often not the case. As indicated by missing data (marked n.a.) in Table 4, we could not access such relevant and necessary information for three of the studies.[3] When a study did not address a particular research question, an effect size was naturally impossible to derive. Such cases are marked n.t. in Table 4.

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

Three studies (Jourdenais, 1998; Kubota, 2000; White, 1998) administered delayed posttests, and *d* values were calculated again by contrasting experimental groups with comparison groups on delayed posttest scores. Because effect sizes could be calculated on every possible contrast depending on the number of independent and dependent variables, each primary study often produced more than two effect sizes. As noted by Norris and Ortega (2000), multiple effect sizes from one study might be averaged for an examination of the combined impact of the study's treatment across tasks. For this meta-analysis, however, if a primary study investigated different target forms with different pools of participants—that is, if different groups of participants experienced different targets—each group was considered an independent study sample. This choice was reasonable not only because target structure might be an important moderating variable (Norris & Ortega) but also because it is plausible that the distinctive characteristics of participants across studies justify our treating each study as an independent study sample. This reasoning resulted in three *d* values for Jourdenais and two *d* values each for Kubota and Overstreet (2002) (cf. Table 4).

Calculated effect sizes were interpreted following Cohen's (1988) suggestions for the characterization of the effect size values: *d* values around 0.20 should be understood as small effects, 0.50 as medium effects, and 0.80 and above as large effects. To verify statistical trustworthiness of the produced effect sizes, 95% confidence intervals were calculated and are provided in Table 4. Simply put, two nonoverlapping confidence intervals indicate that the difference of the two averaged effect sizes is statistically trustworthy, whereas overlaps of the two intervals mean nontrustworthiness of observed difference of the averaged effect sizes. Additionally, effect sizes with confidence intervals whose lower boundary touches or contains zero indicate no statistically significant effect.

As shown in Table 4, students who experienced VIE conditions outperformed the comparison groups by 0.22 standard deviation units. In other words, VIE had very small-sized overall effects on learners' grammar learning. The mean ($d = 0.22$) is indeed clearly small when compared with what Norris and Ortega (2000) reported for the overall effects of L2 instruction ($d = 0.96$) as well as the effects of implicit focus-on-form treatments ($d = 0.69$). Additionally, because the confidence interval (−0.03 to 0.47) includes zero, the observed effects are not statistically trustworthy. The impact of VIE on delayed posttests was found to be negligible ($d = -0.13$), as the *d* value was such a small portion of a standard deviation unit. It should be noted again that all of these effect sizes—except those from Jourdenais (1998)—were derived from the comparison of VIE treatments with input flood treatments; that is, the *d* values reflect the effectiveness of VIE over input flood treatments rather than the absolute magnitude of effectiveness of VIE, which could only be measured by comparing VIE to no exposure conditions.

Pre-to-post effect size *d* values were also calculated to investigate the degree to which learners improved after the VIE treatment sessions from the differ-

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

ent perspective of within-group change. Table 4 illustrates that students provided with enhanced versions of texts improved from the pretest to the posttest by a medium-sized effect ($d = 0.55$). This was a considerably smaller effect than that found by Norris and Ortega (2000) for the pre-to-post change associated with overall L2 instruction ($d = 1.66$) and the pre-to-post change resulting from implicit focus-on-form treatments ($d = 1.51$).

To address the third research question, it was necessary to calculate effect sizes on dependent measures of meaning comprehension. Six primary studies (seven unique study samples) contributed to these calculations (see Table 4). The calculations were again done by contrasting the means of the experimental groups with those of the comparison groups on posttreatment reading comprehension measures. The averaged effect size for VIE students' meaning comprehension was small but negative in its direction ($d = -0.26$). Although the confidence interval includes a zero value (−0.70 to 0.19), the pattern of results is suggestive of a tendency for VIE to affect learners' meaning processing negatively.

Research question 4 asked whether publication bias would be at work in this research domain. To address this question, we further analyzed the relative effects of VIE for studies grouped in terms of type of publication: published, published based on dissertation research, and unpublished dissertations. Table 5 demonstrates that the averaged effect sizes for learners' grammar learning on the immediate posttests were $d = 0.55$ for published studies, $d = 0.24$ for published studies based on dissertation research, and $d = -0.01$ for unpublished dissertations. Due to the small number of study samples ($k = 5$, 4, and 8), the difference of the observed averaged effect sizes was not statistically trustworthy, as evidenced by the overlaps of the 95% confidence intervals (−0.37 to 1.47, −0.05 to 0.53, and −0.24 to 0.22, respectively). Additionally, each confidence interval includes zero. Even though the inclusion of more studies would lead us to more solid conclusions, the current findings suggest a publication-bias trend.

This trend was also supported by results from meaning comprehension measures, although in the opposite direction: The averaged negative effect size of the published studies ($d = -0.42$) was larger than that of the unpublished dissertations ($d = -0.24$); there was only one study in the published dissertation category, so the average calculation of effect sizes was not feasible. In both cases, 95% confidence intervals were wide in range (−1.28 to 0.44 and −1.57 to 1.1) due to the small number of study samples ($k = 4$ and 2). Figure 4 graphically shows the publication findings.

DISCUSSION AND CONCLUSIONS

The primary goal of this study was to address the magnitude of the overall impact of VIE on grammar learning. It was found that learners exposed to enhanced texts outperformed learners who read unenhanced texts by a very

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

Table 5. Average effect sizes by publication type

Statistic	Immediate posttest	Delayed posttest	Pre-to-post contrast	Meaning comprehension
	A. Published studies (n = 8)			
k	5	—	3	4
Average effect size	0.55	—	0.68	−0.42
95% CI upper	1.47	—	2.02	0.44
95% CI lower	−0.37	—	−0.66	−1.28
SD	0.74	—	0.54	0.54
	B. Published dissertation articles (n = 4)			
k	4	1	4	1
Average effect size	0.24	—	0.99	—
95% CI upper	0.53	—	1.96	—
95% CI lower	−0.05	—	0.02	—
SD	0.18	—	0.61	—
	C. Unpublished dissertations (n = 4)			
k	8	5	4	2
Average effect size	−0.01	−0.15	0.01	−0.24
95% CI upper	0.22	0.10	0.31	1.10
95% CI lower	−0.24	−0.40	−0.29	−1.57
SD	0.28	0.20	0.19	0.15

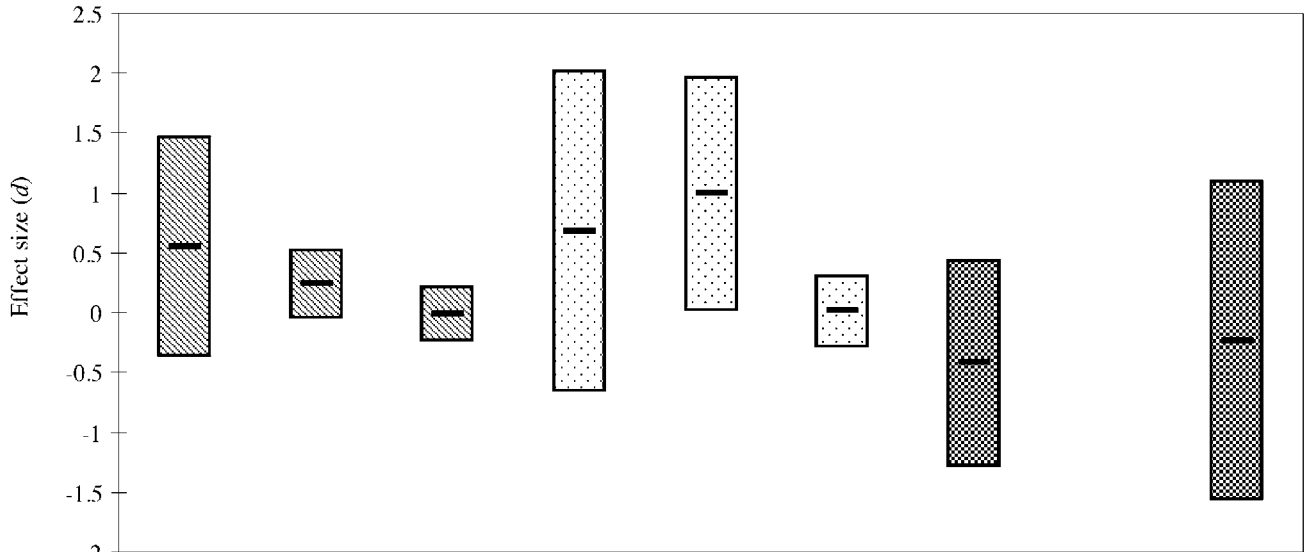
Note. Published studies: Alanen (1995), Jourdenais et al. (1995), Lee (2007), Leow (1997, 2001), Leow et al. (2003), Overstreet (1998), Shook (1994); published dissertation articles: Doughty (1991), Izumi (2002), White (1998), Wong (2003); unpublished dissertations: Ha (2005), Jourdenais (1998; 3 samples), Kubota (2000; 2 samples), Overstreet (2002; 2 samples).

small-sized effect (d = 0.22). However, as already noted, the effect size d values were calculated by contrasting experimental groups with comparison groups rather than true control groups. The comparison groups were provided with another intervention: input flood. For this reason, the small magnitude of the overall effect of VIE should be interpreted as the relative impact of enhancement over flood of input, which in itself is thought to have small but detectable effects on grammar learning (see Trahey & White, 1993, and Williams & Evans, 1998, for discussion on the beneficial effects of input flood). The relatively small-sized effect (d = 0.22) of VIE on learners' grammar learning can only be put into perspective in this relative light.

It was also noteworthy that VIE might have had negative effects on learners' meaning processing, as shown by the small but negative effect size value yielded by the reading comprehension measures (d = −0.26). This could be accounted for by a competition between learners' form and meaning processing. As proposed by a group of primary studies (e.g., Bransdorfer, 1991; Lee, 2007; VanPatten, 1990; Wong, 2001), learners might have difficulty simultaneously paying attention to both form and meaning. More studies in the future

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press



	Grammar learning						Meaning comprehension		
	Group contrast			Pre-to-post contrast					
	Published studies (k = 5)	Published dissertations (k = 4)	Unpublished dissertations (k = 8)	Published studies (k = 3)	Published dissertations (k = 4)	Unpublished dissertations (k = 4)	Published studies (k = 4)	Published dissertations (k = 1)	Unpublished dissertations (k = 2)
Upper 95%	1.47	0.53	0.22	2.02	1.96	0.31	0.44		1.1
Mean	0.55	0.24	−0.01	0.68	0.99	0.01	−0.42		−0.24
Lower 95%	−0.37	−0.05	−0.24	−0.66	0.02	−0.29	−1.28		−1.57

Figure 4. Average effect sizes and 95% confidence intervals of each contrast for the three types of publication.

should address, by design, this theoretical relationship between form and meaning in VIE treatments.

To date, primary researchers have drawn conflicting conclusions on the effectiveness of VIE on grammar learning, and the meta-analytic data presented here confirm that the results are inconclusive in the extant research. We believe the findings presented in our synthesis indicate the divergence of methodological features of the primary study groups as one of the factors that generate such undeterminacy of results. When the 16 primary studies and 20 unique samples identified in the literature search were systematically coded for study features, it became clear that the 16-study pool is indeed heterogeneous in many ways. The 20 unique study samples had distinctive profiles with regard to the background of the study participants, materials used, specific procedures undertaken, and the like. Moreover, the domain of research on the effects of VIE on grammar learning is not yet a fully developed area of inquiry, which is evident from the fact that only four unpublished dissertations and 12 articles published in refereed journals have appeared in the last 15 years or so. Given that several of these studies were produced by the same researchers, it also appears that VIE research has mainly been undertaken within a relatively small research community to date. More studies on VIE are thus needed, because a larger empirical basis would enable us to clarify the research domain with more confidence. We would like to conclude with some reflections on key issues that might contribute to better understanding and eventually improving the research domain of VIE investigations in the future.

Our first recommendation for future studies pertains to the variable of learner proficiency. It was found that researchers of VIE have exclusively relied on participants' institutional status to classify their proficiency levels. This was strikingly different from the observation by Thomas (2006), who, from her analysis of two corpora compiled from articles from four journals (*Applied Linguistics*, *Language Learning*, *Second Language Research*, and *Studies in Second Language Acquisition*) in two time periods, 1988–1992 and 2000–2004, demonstrated that although 40.1% and 33.2% of the studies reported institutional status in the two time periods, respectively, other assessment measures (i.e., in-house assessment and standardized test) have also been used fairly often in the overall field of SLA. Learner proficiency might be one of the decisive factors affecting the ability of learners to focus on form on their own or to benefit from the impact of pedagogic focus-on-form interventions (J. Williams, 1999). Therefore, future studies should deliberately employ additional measures or indexes to accurately gauge the participants' level of proficiency.

A majority of the primary studies recruited participants with minimal knowledge of the targeted forms (n = 13, or 81%). It has not yet been determined how much prior knowledge learners should have in order to benefit from VIE. As might be the case for any intervention technique, VIE would be expected to produce more positive results when learners have some prior exposure to the targeted forms (Lee, 2007). In this regard, it should be noted that two studies that recruited participants with a comparatively high degree of knowledge

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

of the targets (Ha, 2005; Jourdenais et al., 1995) reported contradictory results in terms of the magnitude of effects of VIE on grammar learning. The magnitude reported by Ha was only negligible (d = 0.07), but the averaged effect size for Jourdenais et al.'s study was large (d = 1.62). Prior knowledge of target forms might interact with the nature of the targets, producing these incongruous learning outcomes. For instance, Ha's study targeted determiners, a highly frequent structure in learners' previous experience, and a flood or a VIE treatment might thus have made little contribution to their learning of this structure. Future investigations should be designed to address the interplay between learners' prior knowledge and the nature of target forms, so as to gain answers to the question of appropriate knowledge prior to VIE interventions. Additionally, researchers as well as teachers might need to consider whether L2 readers are developmentally ready to acquire particular target forms, as recommended by a number of researchers (Lightbown, 1998; Mackey, 1999; Mackey & Philp, 1998; Philp, 2003; Spada & Lightbown, 1999).

It was also observed that the VIE studies overall have been short term in nature, with 38% of them (n = 6 of 16) involving only a single treatment session that featured less than 30 min exposure to target forms. Additionally, because there have been only three studies that administered delayed posttests (19%), it was not possible to determine how durable the effects of VIE might be on learners' developing L2 systems, although this could be an important consideration for language pedagogy. Future studies to investigate questions surrounding the appropriate intensity of VIE treatments and durability of VIE benefits would be a welcome addition to the research domain.

It should be acknowledged that the features mentioned so far (e.g., learner proficiency, prior knowledge of the participants, extent of the treatment intensity in light of the number of sessions and length of reading texts, and reported degree of noticing) could be intervening variables that deserve additional investigation via meta-analysis. Yet, in the current study, it was decided not to probe those variables for further analyses, largely due to the small number of study samples. Questions of the relative effects of VIE contingent on these variables, however, would be viable for future studies, extending the inquiries to probe the underlying factors that might be implicated in the process of reading enhancement-embedded texts.

A possible publication bias was noted in the current study. Effect size d values for the immediately administered posttest measures for grammar learning ranged from medium (d = 0.55) to small-sized (d = 0.24) for the group of published studies, whereas the effect size value was negligible (d = −0.01) for the unpublished dissertation studies. Therefore, caution should be exercised in interpreting findings reported by a number of studies published in refereed journals, in that they might not be representative of the true effects of VIE across educational contexts.

The conclusions that can be drawn from the present synthesis are tentative rather than definitive, due in large part to the small number of studies that were viable for the current review, their divergent methodological fea-

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

tures, and the evidence suggesting a publication bias in the research domain. Future cumulative research with explicit reporting—especially on the description of treatment sessions—is sorely needed, so that more reliable empirical findings can better inform teachers regarding their practices with VIE treatment. The present study poses intriguing and valuable research questions worth examining in future research. It is now time to embark on studies that probe the underlying questions of what exactly influences learners' perception of enhanced forms and how the processing of these enhanced forms might facilitate L2 grammar learning.

(*Received 19 March 2007*)

NOTES

1. Sharwood Smith (1981) first discussed theoretical and pedagogical considerations related to VIE with the term *consciousness raising*.

2. We followed Norris and Ortega's (2000) formulas for the calculation of effect sizes and confidence intervals.

3. The three studies did not report a *t* or an *F* value, which would have allowed us to calculate *d* values in the absence of descriptive statistics.

REFERENCES

*Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259–302). Honolulu: University of Hawai'i Second Language Teaching & Curriculum Center.

Barcroft, J. (2003). Distinctiveness and bidirectional effects in input enhancement for vocabulary learning. *Applied Language Learning*, *13*, 133–159.

Bransdorfer, R. (1991). *Communicative value and linguistic knowledge in second language acquisition oral input processing*. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.

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

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.

Doughty, C. J. (1988). *The effects of instruction on the acquisition of relativization in English as a second language*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.

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

Doughty, C. J., & Williams, J. (1998). *Focus on form in classroom second language acquisition*. New York: Cambridge University Press.

*Ha, J. (2005). *Developing English determiners through Internet chat: An experiment with Korean EFL students*. Unpublished doctoral dissertation, University of Florida, Gainesville.

Izumi, S. (2000). *Promoting noticing and SLA: An empirical study of the effects of output and input enhancement on ESL relativization*. Unpublished doctoral dissertation, Georgetown University, Washington, DC.

*Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study of ESL relativization. *Studies in Second Language Acquisition*, *24*, 541–577.

Izumi, S. (2003). Visual input enhancement as focus on form. *Sophia Linguistica*, *13*, 1–30.

*Jourdenais, R. (1998). *The effects of textual enhancement on the acquisition of the Spanish preterit and imperfect*. Unpublished doctoral dissertation, Georgetown University, Washington, DC.

*Jourdenais, R., Ota, M., Stauffer, S., Boyson, B., & Doughty, C. J. (1995). Does textual enhancement promote noticing? A think-aloud protocol analysis. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 183–216). Honolulu: University of Hawai'i Second Language Teaching & Curriculum Center.

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

Kim, Y. (2003). *Effects of input elaboration and enhancement on second language vocabulary acquisition through reading by Korean learners of English*. Unpublished doctoral dissertation, University of Hawai'i, Honolulu.
*Kubota, S. (2000). *Input enhancement in classroom second language acquisition of Japanese*. Unpublished doctoral dissertation, Purdue University, West Lafayette, IN.
*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.
Leeman, J., Arteagoitia, I., Fridman, B., & Doughty, C. J. (1995). Integrating attention to form with meaning: Focus on form in content-based Spanish instruction. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 217–258). Honolulu: University of Hawai'i Second Language Teaching & Curriculum Center.
*Leow, R. (1997). The effects of input enhancement and text length on adult L2 readers' comprehension and intake in second language acquisition. *Applied Language Learning*, *8*, 151–182.
*Leow, R. (2001). Do learners notice enhanced forms while interacting with the L2? An online and offline study of the role of written input enhancement in L2 reading. *Hispania*, *84*, 496–509.
*Leow, R., Egi, T., Nuevo, A., & Tsai, Y. (2003). The roles of textual enhancement and type of linguistic item in adult L2 learners' comprehension and intake. *Applied Language Learning*, *13*, 1–16.
Light, R., & Pillemer, D. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.
Lightbown, P. M. (1998). The importance of timing in focus on form. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 177–196). New York: Cambridge University Press.
Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot, R. B. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). Amsterdam: Benjamins.
Lorch, R. F., Jr., Lorch, E. P., & Klusewitz, M. A. (1995). Effects of typographical cues on reading and recall of text. *Contemporary Educational Psychology*, *20*, 51–64.
Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, *21*, 557–587.
Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal*, *82*, 338–356.
Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.
Norris, J. M., & Ortega, L. (2006). The value and practice of research synthesis for language learning and teaching. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 3–50). Amsterdam: Benjamins.
*Overstreet, M. (1998). Text enhancement and content familiarity: The focus of learner attention. *Spanish Applied Linguistics*, *2*, 229–258.
*Overstreet, M. (2002). *The effects of textual enhancement on second language learner reading comprehension and form recognition*. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.
Pacheco, J. (2004). *A comparison of textual input enhancement and explicit rule presentation in secondary one English as a second language classes*. Unpublished master's thesis, Concordia University, Montreal, Canada.
Philp, J. (2003). Constraints on 'noticing the gap': Nonnative speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition*, *25*, 99–126.
Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition*, *19*, 223–247.
Rosenthal, R. (1994). Parametric measures of effect sizes. In H. Cooper & L. Hedges (Eds.), *Handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.
Sato, K. (2005). *Does instruction help learners become proficient in L2 writing? The case of the Japanese particles wa, ga, and the passive*. Unpublished doctoral dissertation, University of Oregon, Eugene.
Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). New York: Cambridge University Press.
Seiba, Z. (2001). *Classroom instruction and second language acquisition: The effect of explicit form-focused instruction on L2 learners' linguistic competence*. Unpublished doctoral dissertation, University of Mississippi, Oxford.

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

Sharwood Smith, M. (1981). Consciousness-raising and the second language learner. *Applied Linguistics*, *2*, 159–168.
*Shook, D. J. (1994). FL/L2 reading, grammatical information, and the input-to-intake phenomenon. *Applied Language Learning*, *5*, 57–93.
Shook, D. J. (1999). What foreign language reading recalls reveal about the input-to-intake phenomenon. *Applied Language Learning*, *10*, 39–76.
Spada, N., & Lightbown, P. M. (1999). Instruction, first language influence, and developmental readiness in second language acquisition. *Modern Language Journal*, *83*, 1–22.
Thomas, M. (2006). Research synthesis and historiography: The case of assessment of second language proficiency. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 279–298). Amsterdam: Benjamins.
Trahey, M., & White, L. (1993). Positive evidence and pre-emption in the second language classroom. *Studies in Second Language Acquisition*, *15*, 181–204.
VanPatten, B. (1990). Attending to content and form in the input: An experiment in consciousness. *Studies in Second Language Acquisition*, *12*, 287–301.
White, J. (1996). *An input enhancement study with ESL children: Effects on the acquisition of possessive determiners*. Unpublished doctoral dissertation, McGill University, Montreal, Canada.
*White, J. (1998). Getting the learners' attention: A typographical input enhancement study. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 85–113). New York: Cambridge University Press.
Wijaya, J. (2000, May). *Typographical input enhancement for learning Indonesian transitivity*. Paper presented at the tenth annual meeting of the Southwest Asian Linguistics Society, Tempe, Arizona.
Williams, J. (1999). Learner-generated attention to form. *Language Learning*, *49*, 583–625.
Williams, J., & Evans, J. (1998). What kind of focus and on which forms? In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 139–155). New York: Cambridge University Press.
Williams, J. N. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition*, *21*, 1–48.
Witten, C. M. (2002). *The effects of input enhancement and interactive video viewing on the development of pragmatic awareness and use in the beginning Spanish L2 classroom*. Unpublished doctoral dissertation, University of Texas, Austin.
Wong, W. (2000). *The effects of textual enhancement and simplified input on L2 comprehension and acquisition of non-meaningful grammatical form*. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.
Wong, W. (2001). Modality and attention to meaning and form in the input. *Studies in Second Language Acquisition*, *23*, 345–368.
*Wong, W. (2003). Textual enhancement and simplified input: Effects on L2 comprehension and acquisition of non-meaningful grammatical form. *Applied Language Learning*, *13*, 17–45.
Wong, W. (2005). *Input enhancement: From theory and research to the classroom*. Boston: McGraw-Hill.

APPENDIX

CODED VARIABLES

Learner Characteristics

I. Participants
  A. *n* size
    1. Total *n*
    2. *n* per cell
  B. How proficiency levels were defined
    1. Impressionistic judgment
    2. Institutional status

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

3. In-house assessment
4. Standardized test

C. Proficiency levels of participants
1. Beginning
2. Intermediate
3. Advanced

D. Knowledge of the targeted form(s)
1. No knowledge (No formal exposure)
2. Minimal knowledge
3. Some knowledge

E. Length of L2 study
F. Learning context

Research Design

I. Target language
II. Target forms
A. Target forms
B. Number of types
C. Number of tokens
D. Typographical cues

III. Materials
A. Text type
B. Text length (total number of words used for the treatment[s])
1. 0–300 words
2. 301–600 words
3. 601–900 words
4. 901–1,200 words
5. 1,201 and above words

IV. Other independent variables in addition to input enhancement
V. Dependent variables
A. Types of measures used for pretests
B. Measures used for form learning
1. Receptive
2. Productive
i. Free constructed
ii. Constrained constructed

C. Measures used for meaning comprehension, if any
1. Receptive
2. Productive
i. Free constructed
ii. Constrained constructed

D. Measures used for noticing, if any

VI. Treatment intensity
A. Number of treatment sessions
B. Interval between the pretest and the first treatment
1. Both on the same day
2. 1–2 days
3. 3–4 days
4. More than 5 days

C. Interval between the last treatment and the posttest
1. Both on the same day
2. 1–2 days

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press

3. 3–4 days
4. More than 5 days

D. Duration of the treatment(s)
1. 1 day
2. 2 days to less than 1 week
3. 1 week to less than 2 weeks
4. 2 weeks to less than 3 weeks
5. More than 3 weeks

E. Total amount of treatment time
1. Less than 30 min
2. 30 min to less than 60 min
3. 1 h to less than 2 h
4. 3 h to less than 4 h
5. More than 4 h

F. Total amount of reading time
1. Less than 30 min
2. 30 min to less than 60 min
3. 1 h to less than 2 h
4. 3 h to less than 4 h
5. More than 4 h

https://doi.org/10.1017/S0272263108080479 Published online by Cambridge University Press