

# **Comprehension-based versus production-based instruction: A meta-analysis of comparative studies**

**Natsuko, Shintani, Shaofeng Li and Rod Ellis**

**University of Auckland**

Meta-analyses of different aspects of second language (L2) instruction have become very fashionable. They began with Norris and Ortega's (2000) meta-analysis of form-focused instruction studies, which investigated the relative effectiveness of implicit versus explicit instruction and focus-on-form and focus-on-forms. Subsequently, Norris and Ortega (2006) published an edited collection of meta-analytical studies that addressed such issues as access to Universal Grammar (Dinsmore), task-based interaction (Keck, Ibarra-Shea, Tracy-Ventura and Wa-Mbaleka), corrective feedback, effects of instruction on pragmatic development (Jeon and Tadayoshi-Kaya) and explicit reading strategy training (Stevens and Asher). Corrective feedback has attracted other meta-analysts (Lyster and Saito, 2010; Li, 2010). Mackey and Goo (2007) reported a meta-analysis of interaction research in SLA. Spada and Tomita (2010) examined studies that had investigated the effects of implicit and explicit instruction on simple and complex grammatical forms. However, despite the growing popularity of meta-analysis in SLA research, and also despite the large number of studies that have compared comprehension-based and production-based instruction, no attempt has been made to compare the effects of these two types of instruction on L2 grammar learning meta-analytically.

## **Comprehension-based and production-based instruction**

Traditionally language instruction has emphasized the importance of learning through producing sentences in the target language. This is true of grammar translation (e.g. L1 → L2 translation exercises), of the Audiolingual Method (which sought to develop habits by having learners mechanically produce exemplars of specific structural patterns) and present-practice-produce (PPP), which differs from other 'traditional' methods mainly because it incorporates opportunities for free as well as controlled production in the L2. There are obvious reasons for this emphasis on production. First, the goal of most language courses is to enable students to speak and write in the L2 and it would seem self-evident that this can best be achieved by having students speak and write. Second, teachers frequently evaluate the success of their lessons in terms of student participation and view 'participation' as necessarily involving production. Third, students' production, when erroneous, provides opportunities for corrective feedback and 'errors' are conceived as problems with

production not comprehension. Language pedagogy – and the publishing industry that supports it – is quite clearly centred on teaching learners to produce through production. Interestingly and somewhat surprisingly, this is just as true of courses for complete beginners as it is for more proficient L2 learners.

In the 1960s, however, the emphasis on this pedagogy of production was challenged. The key figure in researching comprehension-based language teaching at this time was James Asher who developed a method called Total Physical Response (TPR). The main features of TPR are (1) to delay production until learners are ready to speak, (2) to maximise exposure to the language by introducing grammatical structures through oral commands and (3) to postpone abstract language until learners have reached an advanced level when they are able to infer meaning from context. It should be noted that Asher's aim was not to develop listening proficiency but rather to use listening comprehension as a means for teaching the linguistic properties (mainly grammatical but also lexical) of the target language. Asher and associates (see Asher (1977) for a review) conducted a number of studies designed to compare the effects on learning of Total Physical Response (TPR) and other methods, in particular the Audiolingual Method. Asher's studies showed that comprehension-based instruction not only developed students' abilities to comprehend the L2 but also to speak it. He also claimed that it resulted in increased motivation, reduced language anxiety and greater likelihood that students would continue with their study of the language. However, there were major problems with the design of his studies (it was the 1960s after all!) and, in fact, we were not able to include any of Asher's studies in our meta-analysis.

The rationale for comprehension-based language teaching received a theoretical boost from a seminal article published by Newmark (1966), who argued that L2 learning would proceed more smoothly if teachers stopped trying to 'interfere' in the learning process. Interference for Newmark meant forcing learners to produce specific target features. A comprehension-based approach was more clearly compatible with a theory of pedagogy that emphasised the need to assist learners to learn 'naturally' (i.e. incidentally). In a later article, Newmark (1981) identified a number of positive features of comprehension-based language teaching: (1) it presents input in a manner that was likely to command the learners' attention, (2) it emphasizes meaning over form, (3) it removes the need for the two aspects of teaching that Newmark was most opposed to – namely, drilling and explicit analysis of language. However, he also identified a number of deficiencies. He saw it as only of value in the beginning stages of learning a language and he was critical of the fact that comprehension-based materials at that time were still typically based on an inventory of grammatical structures and lexical items to be taught.

The growing interest in comprehension-based teaching was clearly reflected in the collection of papers in Winitiz's *The Comprehension Approach to Foreign Language Instruction*. In his introduction to this book, Winitz (1981) articulated the key premises of this approach

1. Language rules are most easily and accurately acquired by inference. The basic data are the sentences of a language ...
  2. Language acquisition is primarily an implicit process because the acquisition of linguistic knowledge is not, for the most part, under the explicit control of conscious awareness of the student.
  3. The rules of language are so complexly interrelated and so sufficiently detailed as to preclude errorless learning without exposure to a large part of the grammar of a language ... language acquisition is viewed as non-linear ...
  4. Comprehension is a teaching routine whereby the student is systematically exposed to the sentences of a target language. Production exercises, grammatical drills, and practice in translation are not generally used as teaching routines, although they may occasionally be used to test comprehension.
  5. Speaking will develop given sufficient comprehension training ...
- (p. xvii-xviii).

It should be noted, however, that the contributors to Winitiz's book varied with regard premise (3). That is, comprehension-based teaching was often used as a methodology for teaching predetermined linguistic items (usually grammatical). The decoupling of comprehension-based language teaching from a traditional, linear syllabus was achieved most convincingly in Krashen and Terrell's (1983) *Natural Approach*. In accordance with Krashen's views about L2 acquisition (see Krashen, 1981), the aim of this approach was to supply learners with plentiful comprehensible input that would enable them to learn implicitly and incidentally. Essential tenets of the *Natural Approach* were that the goal of language teaching was to develop communication skills and that this could be achieved by recognizing that comprehension always precedes production, that ability to produce in an L2 emerges only after learners have acquired some language through comprehending input, that 'acquisition activities' are central, and that classroom activities must ensure a low 'affective filter'. The *Natural Approach* did not exclude production activities but it only employed comprehension-based activities in the early stages and continued to emphasise such activities throughout. An important point to recognize about the *Natural Approach*, however, is that production was possible in the early stages but on an entirely voluntary basis. This is a point we return to below.

The 1990s, however, saw the comprehension approach once again used to teach specific linguistic (grammatical) features albeit from a very different theoretical base from the one that early advocates of this approach had rejected. Schmidt (1990; 1993) convincingly argued – contra Krashen – that acquisition was not an entirely implicit process. Learners, he claimed, need to ‘notice’ exemplars of features in the input for acquisition to take place. VanPatten (1990) reported a study that showed that meaning and form compete for learners’ attention and that when learners prioritize meaning they are unlikely to notice specific forms in the input. In his 1996 book he outlined a number of ‘processing principles’ to explain learners’ processing priorities and then went on to outline a form of comprehension-based instruction called ‘Processing Instruction’, which aimed to help learners overcome these natural ways of processing input by directing their conscious attention to those grammatical forms that they typically overlook. Like earlier advocates of the comprehension approach, VanPatten argued that acquisition was input- rather than output-driven. Unlike Winitz and Krashen, however, he rejected the view that learning will always occur ‘naturally’ and automatically if learners are exposed to comprehensible input and argued that what was needed was an approach that focused learners attention on the meanings realised by specific grammatical forms. VanPatten also provided a set of guidelines for designing input-processing teaching materials (e.g. ‘Teach only one thing at a time’ and ‘Learners must do something with the input’). His theory and its application to pedagogy have proved extremely productive. There have been a large number of Processing Instruction studies that have compared the effects of comprehension-based and production-based instruction. These studies are included in the meta-analysis.

However, production-based language teaching was and still is dominant. Of the eight different methods that Richards and Rogers (1996) described, five of them were production-based. The Audiolingual Method relied on oral pattern drills and performance of memorized dialogues. Situational Language Teaching emphasized the importance of the use of context to elicit the guided production of specific grammatical forms. Communicative Language Teaching introduced the use of ‘tasks’ to provide opportunities for free production. The Silent Way was only silent for the teacher: learners were expected to produce. Community Language Learning invited learners to first produce in their L1 and then, with the assistance of the teacher, in the L2. Ur (2011) listed six options for teaching grammar, the first three of which involve production (Options 1: task plus focus on form; Option 2: Grammar explanation plus practice; Option 3: Communication). An inspection of some of the popular grammar books of the last two decades also demonstrates the importance attached to learner production. Ellis (2003) analyzed the learning activities in six grammar practice books and found that all six included copious controlled production exercises while four of them also provided free production activities, In contrast, only two provided any comprehension-based activities.

Teacher guides also tend to emphasize production. Ur (1996), for example describes seven types of ‘practice’ designed to assist the learner to ‘make the “leap” from form-focussed accuracy work to fluent, but acceptable, production’ (p. 83). With the exception of the first type (called ‘Awareness’) all the types involve production of one kind of another.

As Ur made clear, production-based language teaching can involve different types of learner production. Ellis (1997) suggested that these can be categorized in terms of ‘error-inducing’ or ‘error-avoiding’ activities. The former entail a ‘garden path approach’; that is, learners are led to make specific errors of the kind that occur naturally in L2 acquisition and then receive explicit corrective feedback. Tomasello and Herron (1988; 1989) reported studies that suggest that this approach is more effective than traditional error-avoiding instruction. In general, however, production-based grammar teaching has made use of error-avoiding production activities. Ellis distinguished two broad types – text-manipulation and text-creation. The former supply learners with sentences illustrating the target structure and require them to operate on these in some way (e.g. by substituting one grammatical form for another, filling in a blank, or transforming sentences from one pattern to another). Text creation activities require students to compose their own sentences. Finer distinctions can also be made. For example, drawing on Ur’s types of practice, text-manipulation activities can consist of ‘controlled drills’, ‘meaningful drills’ and ‘guided, meaningful practice’ while text-creation activities can involve ‘free sentence composition’, ‘structure-based discourse composition’ or ‘free discourse’. In the meta-analysis, however, we have elected to distinguish production-based teaching only in terms of ‘text-manipulation’ and ‘text-creation’ activities as reports of studies often do not provide sufficient detail about the instructional activities employed to make finer distinctions possible.

There are also theoretical reasons for distinguishing text-manipulation and text-creation activities. Behaviourist theories of learning underpinning the Audiolingual Method emphasise the importance of mechanical drilling to build ‘habits’. Skill-learning theory is premised on the assumption that learning grammar involves the transformation of declarative into procedural knowledge through practice consisting of first text-manipulation activities and then text-creation activities. Theories of implicit learning (N. Ellis, 2005; Long and Robinson, 1998) emphasise the importance of free production supported by ‘focus on form’ (i.e. drawing learners’ attention to linguistic problems that arise in their communicative production) as a means of developing their implicit knowledge. Current research (e.g. Ellis, Loewen and Erlam, 2006; Lyster, 2004) investigating the effects of corrective feedback on learner errors in the context of task-based instruction has adopted this perspective. In contrast, text-manipulation activities are seen as more likely to result in explicit than implicit knowledge. Such a position accords with current views about *transfer appropriate learning* which

claims that ‘we can better remember what we have learned if the cognitive processes that are active during learning are similar to those that are active during retrieval’ (Lightbown, 2008: 27). In other words, cognitive theories emphasise that different kinds of production practice foster different kinds of learning and different learning outcomes. Thus, text-manipulation activities will result in the kind of learning that will only be evident in ‘constrained constructed responses’ whereas text-creation activities are more likely to result in learning evident in ‘free constructed responses’.

### **Distinguishing comprehension-based and production-based instruction**

The essential difference between comprehension-based and production-based instruction rests on whether production is or is not required. Comprehension-based activities are designed to provide learners with input only. However, they do not proscribe production; learners are free to engage in both social and private speech when responding to the input. Production-based activities also provide learners with input but they also require a response that involves production. Comprehension-based instruction can be implemented solely through one-way interaction (i.e. teacher to students) or it can also involve two-way interaction when students elect to respond verbally to the input. Production-based instruction invariably involves two-way interaction between teacher and students or between students working in pairs or groups.

However, although it is possible to identify a clear difference between comprehension-based and production-based instruction, it is important to acknowledge that neither type of instruction is monolithic; there is considerable variation in both. For example, both types can be ‘focused’ or ‘unfocused’ depending on whether the syllabus on which they are based is synthetic or analytic. Focused instruction involves the pre-selection of target language features. In comprehension-based instruction, these are then embedded in the input in such a way that learners are forced to process them (i.e. establish a form-meaning mapping) in order to comprehend the input. VanPatten (1996) calls this kind of input ‘structured input’. In production-based instruction, production of the target features is elicited by means of either text-manipulation or text-creation activities. It is of course much easier to ensure that learners produce the target feature in text-manipulation activities. As Loschky and Bley-Vroman (1993) pointed out, it is very difficult to design structure-based production tasks that make production of the target feature ‘essential’. In unfocused instruction there is no pre-selection of target features for learning. In both comprehension-based and production-based instruction this means that unfocused tasks (Ellis, 2003) are employed. In the case of comprehension-based instruction, such tasks simply expose learners to general samples of language. In the case of production-based instruction, they create opportunities for learners to use whatever linguistic resources are at their disposal. However, opportunity exists for focus on form when

unfocused tasks are implemented through the feedback that learners receive on their comprehension or production.

Each type of instruction also varies in terms of whether it caters to intentional or incidental language learning. Incidental acquisition is defined as ‘picking up of words and structures simply by engaging in a variety of communicative activities’ (Hulstijn, 2003; 349). That is, learners acquire ‘naturally’ as a result of exposure to the L2 and/ or opportunities to use it. An assumption of instruction directed at incidental learning is that it is meaning- rather than form-based; that is, it directs learners’ primary attention on message content rather than on the linguistic forms that encode this content. However, this does not mean that there is no ‘noticing’ of these linguistic forms. Incidental learning does not imply totally implicit learning. Nor does it imply that the instructional activities must be of the unfocused kind. Focused tasks that do not make the target feature(s) explicit to the learners can also cater to incidental acquisition. Intentional learning involves ‘the deliberate committing to memory of thousands of words ... and dozens of grammar rules’ (Hulstijn, 2003; 349). Thus, when the instruction makes the target features explicit it invites intentional learning. This can be achieved in a variety of ways. The main one is by some type of presentation stage prior to the practice stage of a lesson. In both comprehension-based and production-based instruction, the target features may be explicitly explained to the learners or alternatively they may be provided with examples to ‘prime’ their attention during the practice stage. An important question, which we address in the meta-analysis, is whether the type of learning (incidental vs. intentional) influences the relative effects of comprehension-based and production-based instruction on acquisition.

### **Theoretical perspectives**

It should be clear from the above account of comprehension-based and production-based instruction that both receive theoretical support from SLA. We conclude with a discussion of theoretical perspectives on an issue that involves very different claims about the effectiveness of the two types of instruction. One purpose of our meta-analysis is to test these claims. The issue concerns the skill-specificity of language instruction. Four different positions are possible: (1) comprehension-based instruction benefits receptive but not productive L2 knowledge whereas production-based instruction benefits productive knowledge but not receptive knowledge, (2) production-based instruction will prove superior to comprehension-based instruction in developing both types of knowledge, and (3) comprehension-based instruction will prove superior to production-based instruction in developing both types of knowledge. .

The first position is supported by skill-learning theory. This emphasises that the power law of practice leads to qualitative changes in the learner's knowledge system over time but only 'in the basic cognitive mechanisms used to execute the same task' (DeKeyser, 2007; 99). This is because the kind of knowledge that characterizes the later stages of development (i.e. the automatization stage) is highly specific and so does not transfer to tasks that are dissimilar from those used to develop the knowledge. In other words, the knowledge that results from comprehension-based instruction will be available in receptive tasks while that which results from production-based tasks will be available in productive tasks. A number of studies (e.g. DeKesyer, 1997; DeKeyser and Sokalski, 2001; Tanaka, 2001) lend some support to this claim. However, as we will see, others do not.

The second position receives support from theories that emphasize the importance of production in L2 acquisition. The Output Hypothesis, for example, proposes that advanced acquisition requires opportunities for 'pushed output'. Swain (1985) argued that comprehensible input alone was not sufficient to ensure high levels of linguistic competence. She suggested that production was beneficial when it pushed learners to produce messages that are concise and socially appropriate. As Swain (1995) noted 'learners ... can fake it, so to speak, in comprehension, but they cannot do so in the same way as production (p. 127). Drawing on Swain's and Skehan's (1998) positions, Ellis (2008) suggested that there are two basic ways in which output promotes acquisition: it enables learners to practise what they already know and so helps them to automatize their linguistic knowledge and it requires syntactic rather than just semantic processing. It should be noted, however, that the Output Hypothesis addresses the role played by free, communicative production, not the kind of controlled production resulting from grammar drills and exercises.

The third position is supported by computational models of L2 acquisition that propose a single knowledge store that is drawn on for both comprehension and production and that develops as a result of input. VanPatten (2007) argued that input processing (defined as the process by which a form-meaning connection is established) in conjunction with other cognitive processes (such as 'restructuring') leads to changes in the learner's internal grammar and these will be subsequently manifest in both receptive and productive tasks. In other words, VanPatten disputed that instruction is skill-specific and saw acquisition as primarily input-driven. Several studies (e.g. VanPatten and Cadierno, 1993; Benati, 2005) based on VanPatten's theory lend support to the superiority of comprehension-based instruction over production-based instruction in developing both receptive and productive knowledge. This position can also be seen as supported by Pienemann's claims regarding the learnability and teachability of grammatical structures. Pienemann proposed that the ability to produce a specific grammatical structure depends on whether the learner has mastered the specific



processing operation that is required. The Teachability Hypothesis predicts that ‘instruction can only promote language acquisition if the interlanguage is close to the point when the structure to be taught is acquired in the natural setting so that sufficient processing prerequisites are developed’ (Pienemann, 1985; 37). ‘Instruction’ here refers to production-based activities. However, Pienemann acknowledged that the processes involved in comprehension and production are different. Thus, he saw no problem in exposing learners to grammatical structures that are beyond their current level of development providing they are not required to produce them. One might conclude, therefore, that comprehension-based instruction is potentially less damaging as it avoids the danger of forcing learners to produce a structure they are not ready for, which according to Pienemann (1989) can actually delay L2 development.

The meta-analysis reported below was undertaken to investigate two pedagogic constructs – comprehension-based and production-based instruction – and the various factors that might influence the relative effect that these two types of instruction have on the learning of L2 grammar. However, as the above account of the various theoretical positions that inform these two types of instruction demonstrates, comparing their effects on the learners’ receptive and productive knowledge of grammar also allows for competing theoretical claims to be tested.

## **Research Questions**

1. Is there any difference in the effect of comprehension-based and production-based instruction on the acquisition of receptive knowledge of grammatical features?
2. Is there any difference in the effect of comprehension-based and production-based instruction on the acquisition of productive knowledge of grammatical features?
3. What factors moderate the effects of (a) comprehension-based and (b) production-based instruction on L2 grammar acquisition?

## **Method**

### ***Meta-analysis***

A meta-analysis (Glass, 1976) or “research synthesis” (Cooper & Hedges, 1994) is defined as a quantitative research method which involves combining and synthesising the results of multiple studies to provide a macroscopic view of a particular research domain. Typically in a meta-analysis effect sizes (in the form of standardized mean difference, odd ratio, correlation coefficient, etc.) are calculated (or directly extracted if such numerical indexes are provided) based on the descriptive statistics provided in primary empirical studies. This is then followed by the aggregation of the effect

sizes to arrive at a summative numerical index that represents the magnitude of a certain relationship or the effect of a treatment across studies.

Another important meta-analytic procedure is the identification of factors mediating the effects of the treatment or relationship of the construct under investigation. This is called moderator analysis.

Moderators are the methodological features (e.g., immediate versus delayed effect; incidental vs. intentional learning) that are either reported in the primary studies or extracted by the meta-analyst that hypothetically account for a significant portion of the variance of the obtained set of effect sizes. A primary study in a meta-analysis is analogous to a participant in an empirical study. The participants in each primary study constitute a sample of the involved population.

Norris and Ortega (2006) pointed out, a danger of employing meta-analysis is that it is ‘un-meta-analyzable’ (p.10) because of the lack of empirical research on the same topics. In fact, nearly all published meta-analyses identified the need for more studies in the particular research domain they investigated (e.g. Jeon & Kaya, 2006). However, comparative studies of comprehension-based and production-based instruction have accumulated over years making it feasible to carry out a meta-analysis. It was for this reason that we decided to conduct a meta-analysis to investigate the relative effectiveness of the two types of instruction. The meta-analysis sought answers to the following questions.

### ***Selection of studies***

The following *criteria* were used to select studies for the meta-analysis:

- 1) The studies had to be published between 1991 and April 2011 which is the point that the selection of studies for this analysis was completed. This decision was made because although debate about the relative effectiveness of CBI and PBI started much earlier there were few empirical studies prior to 1991 and those that did exist - such as Asher’s (1977) experiments with total-physical response (TPR) - did not report sufficient data for the calculation of effect sizes. However, studies published in the last 20 years were sufficient for the purpose of this meta-analysis.
- 2) Only those studies that compared a CBI treatment with a PBI treatment were selected. As noted earlier, the fundamental distinction between CBI and PBI is whether or not the instruction requires the production of the target language rather than whether or not production occurs during the instruction. To have included studies that involved either CBI or PBI would have necessitated a huge amount of time, which was not available for this

project. Also including only comparative studies of the two approaches enabled the meta-analysis to focus on a narrow set of variables relating to the two approaches.

- 3) CBI treatment was defined as comprising any activity that required learners to comprehend oral or written L2 input to achieve the goal of the activity, but required no L2 production or only limited production (e.g. 'yes' or 'no'). Production-based instruction was defined as comprising any activity that required learners to produce the specific target feature(s) in either oral or written form.
- 4) Only studies that involved measurements of either grammar acquisition were included. Grammar acquisition was operationalized as L2 morphology, syntax, or pragmalinguistic features (e.g. Takimoto, 2009).
- 5) Only studies that provided enough statistical information for computing effect sizes were included. The detailed criteria for selecting studies that allow for this are considered below.

The current meta-analysis did not include fugitive literature (e.g., unpublished doctoral dissertations and conference presentations). The literature search started with ten major SLA journals: *Studies in Second Language Acquisition*, *Language Learning*, *Applied Linguistics*, *Language Teaching Research*, *System*, *The Modern Language Journal*, *The Canadian Modern Language Review*, *Language Awareness*, *Computer Assisted Language Learning*, *TESOL Quarterly*. Electronic databases such as the Education Resources Information Center (ERIC) and the Linguistic and Language Behaviour Abstracts (LLBA) were used to extend the search. Some of the key words used in the database search were: *input*, *output*, *comprehension*, *production*, *receptive*, *productive*, *comparative*, *comparison*, *classroom*, *vocabulary learning*, and *acquisition*. As a result, some articles from other journals (*Hispania*, *Second Language Research*, *Spanish Applied Linguistics*, *Foreign Language Annals*, *Applied Psycholinguistics*, *JALT journal* and *RELC Journal*) and chapters from some books (Coady & Huckin, 1997; Bogaards & Laufer-Dvorkin, 2004; VanPatten, 2004; Lee & Benati, 2007; Benati & Lee 2008; Benati, 2009) were identified and selected.

When the same experimental study was reported in more than one published article, only one was selected for the meta-analysis. When a single study report presented more than one experiment, each experiment contributed an effect size in the meta-analysis. In other words, each experiment was treated as a separate study. As a result, 35 unique sample studies reported in 30 published studies were included in the meta-analysis. The lists of the published studies are shown in Table 1.

Table 1

## *Studies on Grammar Acquisition for the Current Meta-Analysis*

---

Allen (2000)	Lee and Benati (2007a)
Benati (2005)	Lee and Benati (2007b)
Benati (2009)	Leeser (2008)
Benati and Lee (2008)	Morgan-Short and Bowden (2006)
Benati, Lee, and Houghton (2008)	Nagata (1998a)
Benati, Lee, and Laval (2008)	Nagata (1998b)
Collentine (1998)	Qin (2008)
De Jong (2005)	Shintani and Ellis (2010)
DeKeyser and Sokalski (1996)	Song and Suh (2008)
Erlam (2003)	Takimoto (2009)
Farley (2001a)	Tanaka (2001)
Farley (2001b)	Toth (2006)
Gass and Torres (2005)	Van Patten and Cadierno (1993a)
Izumi (2002)	VanPatten and Wong (2004)
Koyanagi (1999)	VanPatten, Inclezan, Salazar, and Farley (2009)

---

### ***Coding***

A total of 35 unique sample studies were coded for independent, dependent and moderator variables.

#### *Independent variables.*

The independent variables concerned treatment mode (i.e. CBI or PBI). A treatment was coded as 'CBI' if it required the learners to demonstrate comprehension of the target L2 with no requirement to produce the target L2 features, whereas it was coded as 'PBI' if it required oral or written production of the target features. When a treatment required both comprehension and production of the target L2 features, it was coded as 'PBI' on the grounds that any kind of PBI provides opportunity for both input and output.

#### *Dependent variables.*

Dependent variables were effect sizes for the scores of the measurements after the instructional treatment. In order to take the durability of the instructional effects, the tests in each unique sample study were first classified as either an 'immediate post-test' (i.e. it was conducted within 7 days after the treatment) or a 'delayed post-test' (i.e. it was conducted 8 days after the treatment or later). It should be noted that this classification did not always accord with how the tests had been originally labelled in the studies. For example, Song & Suh's (2008) and Takimoto's (2009) so-called

‘immediate post-tests’ were actually conducted more than one week after the treatment was concluded and were thus classified as ‘delayed’ for the purpose of this meta-analysis. The individual tests were further categorized as either ‘receptive’ or ‘productive’. A receptive test refers to a test that requires the learners to comprehend oral or written L2 input but does not require the production of the target features. A productive test refers to a test that requires the learners to produce the target L2 feature in oral or written form.

### *Moderator variables*

The meta-analysis investigated six moderator variables: (1) instructional context, (2) research setting, (3) the participants’ age, (4) intentional versus incidental learning, (5) CBI involving/ not involving processing instruction and (6) PBI involving text-creation/ text-manipulation. The rationale for choosing each moderator variable and the operational definition of them are provided below.

*Instructional context.* VanPatten suggests that the learners of beginner or low intermediate levels have limited processing capacity, and that therefore requiring production affects the process of intake. This suggests that the effect of CBI and PBI might be influenced by the learners’ L2 proficiency. However, it was difficult to determine the learners’ proficiency levels from the reports of the preliminary studies. For example, it is difficult to determine the proficiency level from information such as ‘university EFL students in Japan’ (Izumi, 2002) or ‘second year university students in the Spanish class’ (VanPatten & Cadierno, 1993). Thus we decided to use the instructional context to examine the factorial influence of the learners’ L2 learning experience. In the EFL context, the learners’ exposure to the L2 is generally limited to the classroom while in the ESL context, the learners have opportunity to use the L2 in the daily life. This might influence the effectiveness of CBI or PBI. In this coding, the instructional context was coded as ‘second language context’ and ‘foreign language context’.

*Research setting.* The influence of research settings on interaction or learning outcomes has been controversial. Several studies have found that interaction in a laboratory setting differs from that in a classroom setting. They found that the mean effect sizes of post-tests were significantly larger in studies conducted in laboratory settings than in studies conducted in classrooms. Gass, Mackey, & Feldman (2005; 2011), on the other hand, investigated three aspects of interaction (negotiation of meaning, language-related episodes, and recasts) of Spanish learners while performing different tasks in the two settings. No statistically significant differences were found between the two settings. Mackey and Goo (2007) reported a meta-analysis of task-based studies that examined whether the setting had any effect on the learning that resulted from interactive tasks. They reported that

interactional feedback was found to be more effective in a foreign language context than in a second language context. In this study, thus, it was decided to examine the effects of research setting on the effectiveness of CBI and PBI. The research setting of a study was coded as ‘laboratory’ or ‘classroom’.

*Participants’ age group.* Shintani (2011) found that young learners benefitted from input-based tasks more than production-based activities in vocabulary acquisition. She explained that this might be because of young learners’ ‘fast-mapping’ ability which has also been observed in L2 acquisition in some studies (Ellis & Heimbach, 1997; Rohde & Tiefenthal, 2000). Thus it was decided to examine the participants’ age as one moderator factor. The learners in each study were classified as belonging to one of the following three age groups: ‘children’ (up to 13 years old), ‘adolescents’ (14-17 years old), and ‘adults’ (18 years old or more). The participant information provided in the original study was interpreted if the learners’ age was not specified.

*Intentional/ incidental learning.* This variable was employed to investigate whether the instruction invited intentional learning or incidental learning had any impact on the comparative effects of CBI and PBI. In this study ‘intentional learning’ was defined as instruction that directed learners to learn the target linguistic structure or words and ‘incidental learning’ was defined as instruction aimed at learning while the learner’s attention was focused on some other aspect of the L2. A treatment was coded as ‘intentional’ if the learner received a written or oral explicit explanation of the target feature as part of the instruction. It was also coded as ‘intentional’ if the instruction did not involve explicit instruction but clearly directed the learners’ attention to a particular linguistic form. Two studies (Lee & Benati, 2007b; VanPatten and Cardiero, 1993) were categorized in this way. Although these studies introduced ‘structured input’ without explicit grammar instruction, the structured input activity (i.e. providing a number of relatively short sentences with one specific target structure; no contextual connection between the sentences) clearly invited intentional learning of the target structures. A treatment was coded as ‘incidental’ if a treatment did not involve any explicit instruction of the target feature AND did not direct learners’ attention to a specific target form. Thus, if a treatment provided some explicit information about non-targeted feature in order to help the learners to perform the activity (e.g., providing the L2 words needed to perform the activity), it was coded as ‘incidental’.

*Whether or not the study involved a CBI group using processing instruction (PI).* This category was included because many of the CBI studies selected for this meta-analysis (i.e. 22 studies out of 33) involved ‘processing instruction’ (VanPatten, 1990) which we thought could be a major factor for the efficacy of the instruction. VanPatten argues manipulating input in the way that it requires learners to process a particular grammatical structure fosters the learners to notice and intake the form

and that therefore, facilitates acquisition. The instructional treatment was coded as ‘PI’ when the study claimed that the experimental treatment involved processing instruction and coded as ‘non-PI’ if the study did not state that it did. Allen (2000) argues some of the replicated processing instruction studies (e.g. DeKeyser & Sokalski, 1996) does not operationalized ‘structured input’ as it was defined by VanPattern (p. 72). However, in this study all studies claiming that they used processing instruction were coded as ‘with PI’ in order to avoid the coding highly interpretative.

*Whether or not the study involved a PBI group providing the opportunity for text-creation.* This category was included because, as discussed, both pedagogical and theoretical viewpoints, the distinction of text-creation and text-manipulation has an impact on L2 acquisition. In the coding, the study was coded as text-creation if the PBI treatment required students to compose their own sentences at any stages of the instruction. For example, Toth (2006) was coded as text-creation as the PBI group engaged in ‘communicative output’ tasks involving summarizing comparing and contrasting pictures or narrating personal stories of recent “unplanned occurrences” (p. 339). Although dictogloss tasks are designed to direct the learners’ attention to form by requiring them to reconstruct the sentences accurately, they were coded as text-creation as the task presumably involved some communicative L2 interaction in the process of completing the tasks (e.g. Izumi, 2002; Leaser, 2008; Qin, 2008; VanPatten et al, 2009). On the other hand, the study was coded as text-manipulation if the PBI involved only controlled production of the L2 such as substituting one grammatical form for another, filling in a blank, or transforming sentences from one pattern to another. Although some studies called their PBI ‘meaning-oriented instruction’, many of them reported the instructional procedures and examples of materials which only indicated controlled production of the L2 (e.g. Lee and Benati, 2007). Those studies were coded as ‘text-manipulation’.

Although the last two factors moderate only either type of instruction, we decided to include them as moderator variables because, as explained, we considered them as possible major factors that decide the efficacy of CBI and PBI, and therefore might affect the comparative results.

### ***Data analysis***

Recall that an effect size may take various forms depending upon the construct to be investigated such as standardized mean difference (e.g., Cohen’s *d*), correlation coefficient (*r*), or odds ratio (*OR*—the odds of an event occurring in one condition as compared with another). This meta-analysis examines the effectiveness of two types of instruction, and thus the effect size index relates to mean differences. Standardized mean difference is calculated by dividing the mean difference between two groups by “an estimate of its variation” (e.g., pooled standard deviation) (Cornell & Mulrow, 1999, p.

288). In this study, Cohen's  $d$  was used as the index of effect size estimate, which was interpreted following Cohen's criteria (1988):  $0.80 < d$  constitutes a large effect size,  $0.50 < d < 0.80$  a medium effect size, and  $0.20 < d < 0.50$  a small effect size. Effect sizes were also interpreted with reference to the 95% confidence interval and related  $p$  values. A 95% confidence interval suggests that if the same treatment is repeated many times, 95% of the time the effect would fall within that range. A narrow confidence interval indicates a robust finding. If a confidence interval crosses zero and the  $p$  value is above .05, the null hypothesis that the effect of a treatment or intervention is significantly different from zero is rejected. However, a nonsignificant confidence interval should not be interpreted as lack of any effect: "If CIs also cover effect size (ES) values that are clinically significant, we cannot rule out the possibility that there is a clinically significant average effect that could not be detected in the analysis" (Littell, Corcordan, & Pillai, 2008, p. 135).

As with most other meta-analyses, this meta-analysis follows a two-step procedure: effect size aggregation followed by moderator analysis. Effect size aggregation generates an average effect size that indicates the overall effect of an instruction type or treatment. The purpose of moderator analysis is to identify variables that potentially mediate the effect of the treatment and that may account for the within-subject variability among the effect sizes.  $Q$  tests were performed to determine whether a variable was a significant moderator. All analyses were conducted using Comprehensive Meta-Analysis (Borenstein, Hedges, Higgins, & Rothstein, 2005), professional software developed by a group of leading experts in meta-analysis.

Three categories of effect sizes were computed in this meta-analysis: comparative effect sizes, absolute effect sizes, and within-group effect sizes. A comparative effect size represents the effect of comprehension-based instruction in comparison with production-based instruction. In order to make the results comparable across included studies, data associated with comprehension-based instruction were consistently coded as experimental data, and data pertaining to production-based instruction were coded as baseline data. Thus a significant comparative effect size  $d = .3$  would suggest that comprehension-based instruction is more effective than production-based instruction by .3 standard deviation units. Conversely, a significant effect size in the negative direction  $d = -.3$  would imply that production-based instruction has larger effects than comprehension-based instruction and the effects differ by .3 standard deviation units. An absolute effect size shows the effect of comprehension- or production-based instruction in relation to a control condition. This group of effect sizes were based on the contrasts between comprehension-based instruction or production-based instruction and no instruction or instruction that did not relate to the target structure (control). A within-group effect size reflects the effect of an instruction type as a result of the treatment; it involves the contrast between pretest and posttest scores.



Notwithstanding the fact that three types of effect sizes were computed to obtain an initial, holistic picture of the effects of the two instruction types, it is the comparative effect sizes that were used in subsequent moderator analyses. This decision was based on the following considerations. First, a comparative effect size involves the direct contrast between the two instruction types and provides straightforward and transparent information about the superiority/inferiority of one to the other. Second, using comparative effect sizes made it possible to minimize the statistical analysis needed (i.e. extra group comparisons) and reduce Type I error rate. Absolute or within-group effect sizes are not associated with direct contrasts between comprehension- and production-based instruction, and comparing the two in terms of their effectiveness would involve two steps: aggregating effect sizes related with treatment-control/pre-post contrasts followed by between-group comparisons. Third, analyses based on comparative effect sizes involved all studies included in the meta-analysis whereas analyses based on absolute or within-group effect sizes necessarily precluded studies without a control group or pretest data. Fourth, in the case of absolute effect sizes, the way control conditions were operationalized varied across studies. In some studies, participants in the control group did not receive any treatment; in others they received some placebo treatment that was irrelevant to the target structure; in still others the control group (it is better to be called comparison group in this case) was provided with instruction that was different from comprehension- or production-based instruction. Clearly absolute effect sizes are less accurate indexes of comparative effects.

One issue meta-analysts have to face is sample size inflation, that is, including multiple effect sizes from a single study in the analysis. Lipsey and Wilson (2001) rightly pointed out that the inclusion of more than one effect size in the same analysis violates the assumption of independence of data points and “can render the statistical results highly suspect” (p. 105). To minimize the violation of such assumption and at the same time maintain as much data as possible, a shifting unit of analysis was adopted (Patall, Cooper, & Robinson, 2008). Initially effect sizes from each included study were computed separately. However, when effect sizes were aggregated to obtain overall effects, each study contributed a single effect size, which was the average of several effect sizes if a study generated more than one effect size. When moderator analysis was performed, there was a possibility that more than one effect size was involved in one analysis. However, in calculating the influence of an independent variable, only one effect size was included from each study. For instance, in the event that both a receptive as well as a productive test were used in a study (e.g., Ellis & He, 1999), the two corresponding effect sizes were included in the analysis related to the effect of outcome measure. However, in each of the two categories (receptive and productive), only one effect size was included.

There there are three options for dealing with outliers: (1) including them intact, (2) recoding them to less extreme values (e.g., Winsorizing), and (3) removing them. In this meta-analysis, outliers were

excluded from all analyses in order to ensure the robustness of results. Outliers were detected by examining the standardized values of effect sizes ( $z$  scores): Any value that is more than 2.5 standard deviation units from the mean effect size was considered an outlier. For each set of effect sizes, outlier identification was performed repeatedly until there were no further outliers. It should be noted that outliers are contingent; that is, an outlier in one set of effect sizes is not necessarily an outlier in another set.

A “classical” debate among meta-analysts concerns whether to adopt a fixed-effect model or random-effect model in effect size aggregation (Borenstein, Hedges, Higgins, & Rothstein, 2009; Hunter, 2004; Lipsey & Wilson, 2000; Schulze, 2004). A fixed-effects model assumes that the population mean effect size is the same in all included studies and the variation among effect sizes is only attributable to sampling errors. A random-effect model allows the population effect to vary among studies, and effect size variation results from both within-group and between-group variability. While fixed-effect models were the default, recent meta-analyses (e.g., ) have started to adopt random-effects models because of the potential Type I error rates and misleading narrow confidence intervals associated with the former (Hunter, 2004; Hedges, 1994).  $Q$ -statistic (within-group) is often used to determine which model is appropriate (Hodges, Humphris, & Macfarlane, 2005; Schulz, 2004). A significant  $Q$  indicates heterogeneous distribution and a random-effect model is necessary. In this meta-analysis, a random-effect model was used in effect size aggregation because of the methodological heterogeneity observed among the included studies and the significant  $Q$  values for nearly all aggregated effect sizes including the immediate comparative effects on which all subsequent moderator analyses were based. Moderator analyses were conducted by using an analog to analysis of variance, which “partitions the total homogeneity statistic,  $Q$ , into the portion explained by the categorical variable ( $Q_{\text{between}}$ ) and the residual pooled within groups portion ( $Q_{\text{within}}$ )” (Lipsey & Wilson, 2000, p. 121). A significant  $Q_b$  would mean that the moderator accounts for a significant portion of the variance of the effects of the treatment or intervention.

## **Results**

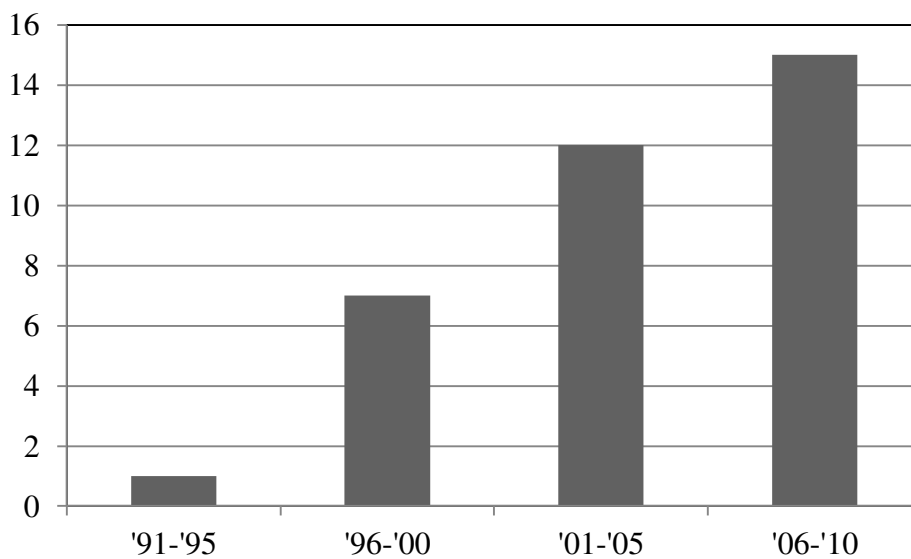
### ***Descriptive results***

A total of 35 comparative studies were retrieved investigating the differential effects of comprehension-based instruction and production-based instruction in second language acquisition. These studies were published between 1991 and 2010, and as shown in Figure 1 there has been a rapid growth in the amount of research on the topic under investigation (from 1 during the first five years to 15 during the last five-year period). Altogether 175 effect sizes were generated from the

included studies, and these effect sizes involved 7,700 codes relating to the dependent and independent variables as well as the methodological features. 93 of the effect sizes were computed based on productive measures, and 82 on receptive measures. 120 represented immediate effects and 55 related to delayed effects.

Figure 1

Publication Frequency of Comparative Studies of CBI versus PBI



A trim-and-fill analysis was performed for the receptive and productive effect sizes (immediate effects) to explore publication bias. It was found that with respect to the receptive effect sizes, two values on the right side of the mean effect were missing. Imputing them would change the mean effect size from  $d = 1.09$ ,  $CI = .64 - 1.55$  to  $d = 1.27$ ,  $CI = .79 - 1.76$ . One hypothetical effect size was imputed on the right side of the mean effect size of the productive effect sizes  $d = -.05$ ,  $CI = -.21 - .11$ , and the adjusted effect size was  $d = -.03$ ,  $CI = -.19 - .13$ . These results demonstrate that in general the extracted effect sizes are normally distributed and that the retrieved studies constitute a reliable representation of comparative studies on the effects of comprehension- and production-based instruction.

Table 2 shows a summary of the methodological features of the included studies. As shown, 32 out of the 35 studies were conducted in foreign language contexts and only 3 in second language contexts. Among the target languages involved, English was the most frequently studied, followed by French, Japanese, and Spanish. In 24 of the studies the participants were undergraduate students

enrolled in university foreign language classes, 6 targeted students at secondary schools, and only 1 investigated elementary school students. 29 studies were conducted in classroom settings and only 6 studies were based in the laboratory. As to the distribution of age groups among the L2 population, 28 investigated adult learners, 6 studied adolescent learners, and 1 examined child learners. 29 involved intentional learning and 6 had to do with incidental learning. Over two-thirds of the studies included processing instruction that is based on VanPatten's Input Processing theory (2007). Learners engaged in text creation in 14 studies and in text manipulation in 24 studies (three studies involved both text creation and text manipulation).

Table 2

## Methodological Features of Included Studies

<i>Aspects</i>	<i>Subcategories</i>	<i>n</i>
Context	Foreign language	32
	Second language	3
Target language	English	9
	French	2
	Japanese	8
	Spanish	16
Instructional status	University	24
	Secondary	6
	Elementary	1
	Miscellaneous	4
Research setting	Class	29
	Laboratory	6
Age	Adult	28
	Adolescent	6
	Child	1
Intentional/incidental	Intentional	29
	Incidental	6
±Processing Instruction	PI	23
	Non-PI	12
Text creation/manipulation	Creation	14
	Manipulation	21

***Meta-Analytic Results***

This section first reports the holistic effects of the two types of instruction in terms of the three categories of effect sizes (i.e. comparative effect sizes, absolute effect sizes and within-group effect sizes) for the ‘receptive’ and ‘productive’ measurements. This is then followed by the analysis of the moderator variables.

## Overall Results

Table 2 summarizes the overall results for absolute effect sizes. Absolute effect sizes for CBI and PBI were calculated by comparing the experimental group with the control group. The results show that both CBI and PBI significantly outperformed the control groups in all the categories with large effect sizes. The  $Q_b$  value showed significant difference between the two groups only for immediate receptive tests ( $Q_b = 5.12, p = .02$ ) indicating that CBI and PBI had difference impact on immediate receptive measurements but not on the other three measurements.

Table 2

### Overall Results: Absolute Effects

Variables	k	Mean $d$	$p$	SE	95% CI		Group Contrast	
					Lower	Upper	$Q_b$	$p$
<i>Absolute Effects</i>								
Receptive								
Immediate								
CB	20	1.96	.00*	.29	1.38	2.54	5.12	.02*
PB	21	1.13	.00*	.22	.69	1.56		
Delayed								
CB	9	1.58	.00*	.29	.99	2.16	1.01	.32
PB	9	1.12	.00*	.34	.45	1.79		
Productive								
Immediate								
CB	22	1.23	.00*	.22	.81	1.66	.07	.79
PB	22	1.32	.00*	.23	.86	1.77		
Delayed								
CB	8	.91	.00*	.19	.54	1.29	.12	.73
PB	8	1.02	.00*	.26	.52	1.53		

Table 3 shows within-group effect sizes. The results showed that both CBI and PBI were significantly effective in both receptive and productive measurement when the scores of pre-tests and post-tests were compared. In all the categories, both groups showed large effect sizes ( $d > 0.8$ ) with significant  $p$  values. In receptive measurements, the  $Q_b$  values showed significant levels ( $Q_b = 21.29, p = .00$ , for immediate tests and  $Q_b = 5.04, p = .02$ , for delayed tests) indicating that CBI and PBI had a significantly different impact on the acquisition of receptive knowledge in both immediate and delayed measurements.

Table 3

Overall Results: Within-Group Effects

Variables	k	Mean <i>d</i>	<i>p</i>	SE	95% CI		Group Contrast	
					Lower	Upper	<i>Q<sub>b</sub></i>	<i>p</i>
<i>Within-Group Effects</i>								
Receptive								
Immediate								
CB	26	2.36	.00*	.26	1.85	2.88	21.09	.00*
PB	27	1.01	.00*	.13	.76	1.27		
Delayed								
CB	13	1.70	.00*	.26	1.19	2.22	5.04	.02*
PB	12	1.02	.00*	.15	.73	1.32		
Productive								
Immediate								
CB	29	1.66	.00*	.19	1.29	2.03	.35	.55
PB	30	1.83	.00*	.22	1.40	2.27		
Delayed								
CB	12	1.23	.00*	.21	.83	1.67	.12	.72
PB	12	1.36	.00*	.22	.93	1.79		

Overall comparative effect sizes were calculated by comparing immediate and delayed post test scores for CBI and PBI. As the ‘immediate measurement’ variable was set to within 7 days after the treatment, two studies were excluded in this analysis; Song & Suh(2008) measured the acquisition one month after the treatment and Takimoto (2009) eight to nine days after the treatment. As for productive tests, immediate tests were conducted in 32 projects and delayed tests were conducted in 15 projects. Table 4 summarizes the results. The positive *d* value for receptive tests indicate that CBI was more effective than PBI and the negative value for productive tests. The *d* value for immediate receptive test is interpreted as large effect size ( $d > 0.8$ ) and the other three are small effect sizes ( $d < 0.2$ ) by Cohen’s criteria (1988). The *p* values for the mean *d* indicate that CBI was significantly more effective than PBI in immediate receptive tests with large effect size ( $d = 1.09, p = .00$ ) but PBI was significantly more effective than CBI with small effect size ( $d = -.21, p = .03$ ) in delayed productive tests. There were no significant differences between the two groups in delayed receptive tests ( $d = .25, p = .11$ ) and immediate productive tests ( $d = -.08, p = .29$ ).

Table 4

## Overall comparative Effects

Variables	k	Mean <i>d</i>	<i>p</i>	SE	95% CI		Heterogeneity	
					Lower	Upper	<i>Q<sub>w</sub></i>	<i>p</i>
Receptive								
Immediate	31	1.09	.00*	.23	.64	1.55	323.43	.00*
Delayed	15	.25	.11	.16	-.06	.55	51.24	.00*
Productive								
Immediate	32	-.08	.29	.07	-.23	.07	48.28	.02*
Delayed	15	-.21	.03*	.09	-.39	-.02	19.29	.15

### *Moderator Analysis*

Table 5 summarizes the results of the moderator analysis by displaying the number of unique sample studies, mean effect sizes, standard error, confidence intervals with the *p* value for each moderator variable. *Q<sub>b</sub>* value with the *p* value on the right columns indicate the difference in the effect sizes between the moderator variables in each category.



Table 5

## Moderator Analysis

Variables	k	Mean <i>d</i>	<i>p</i>	SE	95% CI		Group Contrast	
					Lower	Upper	<i>Q<sub>b</sub></i>	<i>p</i>
<i>Research setting</i>								
Receptive							2.16	.14
Classroom	26	1.26	.00*	.26	.74	1.77		
Laboratory	5	.36	.50	.55	-.71	1.44		
Productive							1.49	.22
Classroom	28	-.01	.97	.09	-.19	.18		
Laboratory	5	-.22	.14	.15	-.51	.07		
<i>Intentional/ incidental learning</i>								
Receptive							3.44	.06
Intentional	27	1.24	.00*	.26	.73	1.74		
Incidental	4	.24	.61	.47	-.69	1.17		
Productive							.03	.85
Intentional	27	-.09	.31	.08	-.25	.08		
Incidental	5	-.05	.82	.19	-.44	.34		
<i>CBI ± PI (Processing Instruction)</i>								
Receptive							24.07	.00*
PI studies	22	1.71	.00*	.29	1.12	2.29		
Non-PI studies	9	-.23	.38	.26	-.73	.28		
Productive							.00	.97
PI studies	22	-.06	.47	.09	-.23	.11		
Non-PI studies	11	-.07	.72	.19	-.44	.30		
<i>PBI + text creation/ text manipulation</i>								
Receptive							2.74	.09
Text creation	11	.69	.04*	.35	.02	1.37		
Text manipulation	23	1.48	.00*	.32	.85	2.12		
Productive							1.47	.23
Text creation	13	-.15	.08	.09	-.32	.02		
Text manipulation	23	.03	.81	.12	-.20	.26		

*Research setting.* Nagata (1998a, 1998b) were omitted for this variables as the studies involved computer-based instruction which we considered as neither of the categories in this coding. The results show that for receptive knowledge, CBI was significantly more effective than PBI in classroom setting with large effect size ( $d = 1.26$ ,  $CI = .74 - 1.77$ ,  $p = .00$ ) but not in the laboratory setting. For

productive knowledge, there were no significant differences between CBI and PBI neither in classroom or laboratory settings. The non-significant  $Q_b$  values for receptive knowledge ( $Q_b = 2.16, p = .14$ ) and productive knowledge ( $Q_b = 1.49, p = .22$ ) indicate that the research setting did not have significant impact on the comparative effectiveness of CBI and PBI. However, the positive small effect size ( $d = .36$ ) and the wide 95% CI indicated that it was not significant due to the small sample size with wide variations but the results were favour to CBI. In other words, the indication of the results is that CBI was more effective than PBI on acquisition of receptive knowledge in both classroom and laboratory settings and the research setting did not moderate the comparative effects.

*Intentional/ incidental learning.* For receptive knowledge, CBI was found to be significantly more effective than PBI when the instruction involved intentional learning ( $d = .50, CI = .21 - .80, p = .00$ ) but no significant difference was found for incidental learning. There were not significant differences between CBI and PBI in the effectiveness on productive knowledge. The non significant  $Q_b$  values ( $Q_b = 3.44, p = .06$  for receptive and  $Q_b = .03, p = .85$  for productive tests) indicate that the intentional/ incidental distinction did not have a significant impact on the comparative effects of CBI and PBI on either receptive or productive knowledge. The ‘incidental learning’ category involved small sample size ( $n = 4$ ) and the confidence intervals crossed zero, indicating that the non-significant results might have been due to the small sample size. However, the  $Q_b$  value that approached to significant level ( $Q_b = 3.44, p = .06$ ) indicated that the distinction of intentional and incidental learning influenced the comparative effects of CBI and PBI. In other words, CBI was indicated to be more effective than PBI in intentional learning but not in incidental learning, and this variable (i.e. incidental or intentional) had an influence on the comparative effects of CBI and PBI.

*CBI group involved / did not involve processing instruction.* CBI was significantly more effective than PBI on receptive knowledge with a large effect when the CBI treatment involved processing instruction (PI) ( $d = 1.71, CI = 1.12 - 2.29, p = .00$ ) whereas there was no significant difference between the two types of instruction when the CBI did not involve PI. The  $Q_b$  values showed that the involvement of PI had significant impact on the comparative effectiveness of CBI and PBI on the development of receptive knowledge ( $Q_b = 24.07, p = .00$ ) but not on the productive knowledge ( $Q_b = .00, p = .97$ ).

*PBI involved text-creation/ text-manipulation.* CBI was significantly more effective than PBI for receptive knowledge with a medium effect size when PBI involved text-creation ( $d = .69, CI = .02 - 1.37, p = .04$ ) and with a large effect sizes when PBI involved text-manipulation ( $d = 1.48, CI = .85 - 2.12, p = .00$ ). There was no significant difference between the CBI and the PBI for production knowledge. The different levels of the effect sizes for text-creation (i.e. medium) and text-

manipulation (i.e. large) indicate CBI had greater effectiveness when the PBI involved only text-manipulation than when involved text-creation. However, non-significant  $Q_b$  value indicated that this variable did not moderate the comparative effectiveness of CBI and PBI.

## Discussion

The validity of a meta-analysis rests on the extent to which the objects of the comparison have been effectively defined. This has proved a problem in some of the previously published meta-analyses. For example, the distinction between ‘implicit’ and ‘explicit instruction’ in Norris and Ortega (2000) is by no means clear cut. Nor is the distinction between ‘simple’ and ‘complex’ grammatical structures in Spada and Tomito’s (2010) study straightforward. We begin our discussion, therefore, by asking to what extent the objects of our comparison – comprehension-based instruction (CBI) and production-based instruction (PBI) can be clearly distinguished.

CBI and PBI are best seen as two macro instructional types. We distinguished them by defining PBI as instruction that required learners’ to produce the L2 target features and CBI as instruction that did not require production but that did not prohibit it. Based on these definitions we identified 35 research experiments in 30 published studies that had compared CBI and PBI. However, in the process of the selection, we noted that both CBI and PBI studies varied considerably in how they provided learners opportunities for production in PBI and opportunities for comprehension in CBI. Some studies (e.g. Shintani, 2010) reported that the participants in the CBI group engaged in substantial L2 production without being requested to do so. It is possible that this also occurred in other CBI studies but unfortunately most studies failed to report whether the treatment allowed learners to produce the L2 or how much the participants actually produced the L2. Thus, in most cases it proved impossible to identify the extent to which production figured in CBI.

Similarly, we were not able to determine to what extent the PBI treatments afforded opportunities for learners to comprehend input containing the target features. Many of the PBI treatments provided some L2 input as a stimulus for the production activities. For example, the ‘traditional instruction’ and ‘meaning-oriented output’ in the processing instruction studies (e.g. Benati, 2005; Farley, 2001; Lee & Benati, 2007) typically involved explicit grammar instruction and activities that presented sentences illustrating the target grammatical forms (e.g. “Last week, I (visit) my uncle” Benati, 2005). Some PBI treatments even required the learners to first comprehend input which included the target grammatical features as a prelude to the production activities. For example, the PBI in Collentine (1998) included a pair-work interview task where learners needed to comprehend their partners’ production. Where the PBI involved dictogloss tasks (Qin, 2008; VanPatten, Inclezan, Salazar, &

Farley, 2009), learners were required to comprehend and understand the text before they tried to reconstruct it.

There were other differences in the way in which the CBI and PBI were operationalized. Some CBI studies required learners to distinguish the correct form from incorrect form (e.g. DeKeyser & Sokalski, 1996) whereas the others merely exposed the participants to exemplars of the correct target feature (e.g. Izumi, 2002; De Jong, 2005). In PBI, the types of L2 production activities varied from controlled discrete-point production (i.e text-manipulation) to more free production (i.e. text-creation). Both CBI and PBI varied in terms of whether there was provision of explicit grammar information, the participants' age, their proficiency levels, the instructional context, the research setting, the target language and the length of treatment. It was possible to investigate some of the differences as moderator variables but not all due to the relatively small size sample.

So, we are left with the question: Given the variation in the two types of instruction, is it valid to carry out a comparison of the effects of CBI and PBI on learning? We argue that it is. First, as we have shown, a number of studies have compared the two types of instruction. Second, there is a fundamental difference in an approach to teaching that emphasizes input and comprehension and one that emphasizes production. Third, this difference has been the subject of debate among teacher educators. Fourth, the difference is of considerable theoretical interest to SLA researchers. We will now discuss what answers the results of the metanalysis provide to our three research questions.

Research question 1 asked whether there was any difference in the effect of CBI and PBI on the acquisition of receptive knowledge of grammatical features. Table 6 summarizes the overall effects of the two types of instruction on receptive knowledge.

Table 6

Summary of the overall effects for receptive knowledge

	<i>k</i>	Comparative effects	Effect size*	
Absolute effects (compared with the control group)				
Immediate**	20	<b>CBI &gt; Control</b>	1.96	large
	21	<b>PBI &gt; Control</b>	1.13	large
Delayed	9	<b>CBI &gt; Control</b>	1.58	large
	9	<b>PBI &gt; Control</b>	1.12	large
Within-group effects				
Immediate**	26	<b>CBI &gt; Control</b>	2.36	large
	27	<b>PBI &gt; Control</b>	1.01	large
Delayed**	13	<b>CBI &gt; Control</b>	1.70	large
	12	<b>PBI &gt; Control</b>	1.02	large

CB vs. PB				
Immediate	31	<b>CBI &gt; PBI</b>	1.09	large
Delayed	15	<b>CBI = PBI</b>	.25	

\*The levels of effect sizes were estimated as: small =  $.2 < d < .5$ , medium =  $.5 < d < .8$ , and large =  $.8 < d$ .

\*\*  $Q_b$  value showed a significant difference in the effect sizes between CBI and PBI.

The absolute and within-group effect sizes showed that both CBI and PBI resulted in receptive knowledge in both the immediate tests (i.e. administered within seven days of the instruction) and the delayed tests (i.e. administered after seven days). In both cases the effect sizes were large ( $.8 < d$ ). The results also revealed differences between the CBI and PBI. The CBI produced significantly larger effect sizes than PBI in the immediate measurements. The within-group results showed that the CBI resulted in significantly better learning than the PBI in both the immediate and delayed tests. The comparative effect sizes suggested that CBI had a greater effect on the development of receptive knowledge in the immediate tests than PBI but that this advantage diminished over time.

Research question 2 asked whether there was any difference in the effect of CBI and PBI on the acquisition of productive knowledge. Table 7 summarizes the results for productive knowledge.

Table 7

Summary of the overall effects for productive knowledge

	<i>k</i>	Comparative effects	Effect size*	
Absolute effects (compared with the control group)				
Immediate	22	<b>CBI &gt; Control</b>	1.23	large
	22	<b>PBI &gt; Control</b>	1.32	large
Delayed	8	<b>CBI &gt; Control</b>	.91	large
	8	<b>PBI &gt; Control</b>	1.02	large
Within-group effects				
Immediate	29	<b>CBI &gt; Control</b>	1.66	large
	30	<b>PBI &gt; Control</b>	1.83	large
Delayed	12	<b>CBI &gt; Control</b>	1.23	large
	12	<b>PBI &gt; Control</b>	1.36	large
CB vs. PB				
Immediate	32	<b>CBI = PBI</b>	-.08	
Delayed	15	<b>PBI &gt; CBI</b>	-.21	small

\*The levels of effect sizes were estimated as: small =  $.2 < d < .5$ , medium =  $.5 < d < .8$ , and large =  $.8 < d$ .

The results differ from those for receptive knowledge. Although the absolute and within-group effect sizes show that both CBI and PBI had a positive effect on productive knowledge with large effect sizes, there was no significant difference between the effect sizes of the two types of instruction. The results for the comparative effects, however, show that although both types of instruction was equally effective in the immediate post-tests, PBI became more effective than CBI in the delayed tests. In

other words, when the acquisition was measured within seven days (i.e. ‘immediately’ in this meta-analysis), CBI resulted in higher levels of receptive knowledge than PBI with a large effect size while both types of instruction were equally effective in developing productive knowledge. In the delayed tests, however, there was no difference in the effect of the two types of instruction on receptive knowledge but PBI proved more effective for on productive knowledge (with a small effect size). In short, CBI was found to be advantageous in the short term but PBI led to more durable learning. Thus, there is a need to consider two points: (1) the initial advantage of CBI over PBI, and (2) durable effect of PBI.

The initial advantage found for CBI lends support to VanPatten’s (2004) claims. He suggests that learners’ limited processing capacity impedes the learners ability to convert input to intake and that forcing learners to produce in the L2 can interfere with their capacity to notice or attend to linguistic form. On the other hand, the advantage seen for PBI in the long term might be explained by the deeper processing that production requires as proposed by Izumi (2002

These two apparently conflicting positions can be reconciled if we view the acquisitional process in terms of the computational model of second language acquisition (Ellis, 1997). According to this model, the first step in the process of acquiring a new L2 form involves converting input into ‘intake’ (i.e. attending to the feature in the input and rehearsing it in working memory). The next step involves ‘integration’ (i.e. incorporating the feature into long-term memory) and ‘accommodation’ (VanPatten, 1996) (i.e. restructuring of the interlanguage system to accommodate the new feature). The final step is output where the learner accesses the new feature and uses it in speech or writing, a process that can also contribute to acquisition (see Swain, 1995). The two types of instruction differ in terms of the particular step they assist. That is, CBI caters primarily to the initial stage of acquisition (i.e. converting input to intake) while PBI focuses on the process of accessing partially acquired knowledge. Thus, CBI has an initial advantage in the case of receptive knowledge because it induces noticing of the new lexical forms and practices their rehearsal in working memory. It also assists production because as VanPatten suggests comprehension and production draw on the same interlanguage system. However, in the case of production the effects of CBI atrophy because it does not involve the kind of ‘generative use’ that Joe (1998) found was important for vocabulary learning and is likely to be equally important for grammar. It is also possible that CBI is more effective for teaching new features whereas PBI is advantageous for developing productive ability of features that have already been partially acquired. Such a view accords with VanPatten’s claims about the relative contributions of the two types of instruction.

Research question 3 addressed the factors that moderated the effects of CBI and PBI on L2 grammar acquisition. Table 8 summarizes the results of moderator analysis. It should be noted that this analysis only focused on the scores obtained from the immediate tests of learning. Two variables moderated the effect of CBI and PBI on receptive knowledge: (1) intentional/ incidental learning and (2) whether the CBI involved Processing Instruction. The results showed that CBI had an advantage when the instruction invited intentional learning but not when it invited incidental learning. Also, it proved beneficial when the CBI involved Processing Instruction but not when it took some other form. The other two variables (i.e. research setting and PBI involving text-creation or text-manipulation) did not moderate the comparative effects. In both cases, CBI was generally more effective than PBI. As regards productive knowledge, none of the variables investigated were found to moderate the comparative effects of CBI and PBI.

Table 8

Moderator analysis

Moderator variables	<i>k</i>	Comparative effects	Effect size*	
<b>Receptive</b>				
<i>Research setting</i>				
Classroom	26	<b>CBI &gt; PBI</b>	1.26	large
Laboratory	5	<b>CBI = PBI</b>	.36	
<i>Intentional/ incidental learning**</i>				
Intentional	27	<b>CBI &gt; PBI</b>	1.24	large
Incidental	4	<b>CBI = PBI</b>	.24	
<i>CBI ± PI (Processing Instruction)**</i>				
PI studies	22	<b>CBI &gt; PBI</b>	1.71	large
Non-PI studies	9	<b>CBI = PBI</b>	-.23	
<i>PBI + text creation/ text manipulation</i>				
Text creation	11	<b>CBI &gt; PBI</b>	-.69	medium
Text manipulation	23	<b>CBI &gt; PBI</b>	-1.48	large
<b>Productive</b>				
<i>Research setting</i>				
Classroom	28	<b>CBI = PBI</b>	-.01	
Laboratory	5	<b>CBI = PBI</b>	-.22	
<i>Intentional/ incidental learning</i>				
Intentional	27	<b>CBI = PBI</b>	-.09	
Incidental	5	<b>CBI = PBI</b>	-.05	
<i>CBI ± PI (Processing Instruction)</i>				
PI studies	22	<b>CBI = PBI</b>	-.06	
Non-PI studies	11	<b>CBI = PBI</b>	-.07	
<i>PBI + text creation/ text manipulation</i>				
Text creation	13	<b>CBI = PBI</b>	-.15	

\*The levels of effect sizes were estimated as: small =  $.2 < d < .5$ , medium =  $.5 < d < .8$ , and large =  $.8 < d$ .

\*\* The  $Q_b$  value indicated that the moderator variable had significant (or approaching to a significant) effects on the comparative effect sizes of CBI and PBI.

The two moderator variables that influenced the comparative effectiveness of CBI and PBI may not have been independent. 22 out of 27 of the intentional learning studies involved processing instruction. Thus, ‘incidental/ intentional’ moderator variable might have been significant because it correlated with the PI variable. In order to examine the interactive effects of the two moderator variables, effect sizes were calculated for studies that fell into these groups: (1) incidental/ - PI, (2) intentional/ + PI, (3) Intentional/ - PI. The results are shown in Table 9. This analysis indicates that when the instruction involves intentional learning, the apparent advantage that CBI has for receptive knowledge is actually due to the fact that so many of the CBI studies involved PI. In CBI involving non-PI studies, it was the PBI that proved more effective for developing receptive knowledge when the instruction required intentional learning.

Table 9

Number of Studies for PI/ Non-PI and Incidental/ Intentional Learning

		PI		Non-PI	
		k	Comparative effect size	k	Comparative effect size
Incidental learning	Receptive	0		4	CBI = PBI ( $d = .24$ )
	Productive	0		5	CBI = PBI ( $d = -.05$ )
Intentional learning	Receptive	22	<b>CBI &gt; PBI</b> ( $d = 1.71$ )*	5	<b>PBI &gt; CBI</b> ( $d = -.59$ )*
	Productive	22	CBI = PBI ( $d = -.06$ )	6	CBI = PBI ( $d = -.08$ )

\*The difference between CBI and PBI was significant.

Then why was CBI more effective than PBI in PI studies?

Another question that needs to be answered is why PBI was actually more effective than CBI when the studies were ‘intentional learning’ and ‘non-PI’.

It might be worth looking at the actual instruction types in the five ‘non-PI’ studies categorized as ‘intentional’ and ‘non-PI’.



Table 10

CBI and PBI treatments and effect sizes for 'Intentional' 'with PI' studies

Study	Effect size ( <i>d</i> )	CBI	PBI
DeKeyser & Sokalski-96 -1	-1.45	1. A handout with grammar explanation was given.	1. A handout with grammar explanation was given.
DeKeyser & Sokalski-96 -2	- .34	2. The learners worked on written discrete-point multiple choice comprehension activities.	2. The learners worked on written text-manipulation activities (modify the given words to correct forms)
Nagata- 98a	- .41	1. Written grammar explanation was given.	1. Written grammar explanation was given.
Nagata- 98b	- .18	2. The learners worked on computer-based comprehension tasks involving both word-level and sentence-level.	2. The learners worked on computer-based production tasks involving both text-manipulation and text-creation.
Tanaka-01	- .77	1. Provided grammar explanation. 2. The learners engaged in multiple-choice comprehension practices (i.e. listening to a sentence and choosing an correct picture).	1. Provided grammar explanation. 2. Composition tasks (i.e. describing the picture in the L2 by using the given L2 word).

It has to be noted that the moderator analysis only involved immediate measurements. Thus the results confirmed that the way how input was provided moderated the effectiveness of CBI while we could not find variables that moderated the effects maybe because PBI had less effects on the initial acquisition.

## Conclusion

The current meta-analysis was motivated by the growing interest in the effectiveness of comprehension-based and production-based instruction and the relationship between receptive/productive learning and receptive productive knowledge.

### Indication to the three positions

The results indicated that CBI was effective for both receptive and productive knowledge and PBI was also effective for both types of knowledge. The absolute and within-group comparisons exclusively showed large effect sizes for CBI and PBI on both receptive and productive knowledge. The results support the above discussed second position (i.e. output is effective for acquisition) and the third position (i.e. input only can promote acquisition) but not the first position that suggests skill

specificity of L2 learning. One might argue, as we have discussed, that it was not clear how much input the PBI participants had or how much opportunity for output the CBI participants had. However, as some studies (e.g. Shintani, 2011) show, the opportunity for L2 output in CBI is generally much less compared to PBI. Thus, the fact that both CBI and PBI had large effects on both receptive and productive knowledge indicates that the skill learning was not only the result on the same mode of learning. Pedagogically, the results suggested that conscious production practice, which PBI generally requires, was not necessary for productive (as well as receptive) grammar acquisition to take place.

This meta-analysis did not involve any ‘unfocused’ (i.e. there is no predetermined specific linguistic features) instruction due to the nature of comparative study. This is probably because of the difficulty of testing as the learning in unfocused instruction only occurs incidentally in the means of negotiation of meaning in CBI and corrective feedback in PBI. The linguistic features cannot be predetermined and thus the measurement in empirically is very difficult. Although studies have intensively examined the effectiveness of both types of focus on form, no comparative studies of CBI and PBI has involved ‘unfocused’ instruction.

## Appendix

Unique sample studies	Year	Context	L2	Inst. Status	Setting	Age	Invited learning	PI/ Non-PI	Text-creation/ manipulation
Allen (2000)	2000	FL	Spanish	Secondary	Class	14-17	Intentional	PI	Manipulation
Benai (2005) - Study 1	2005	FL	English	Secondary	Class	14-17	Intentional	PI	Creation
Benai (2005) - Study 2	2005	FL	English	Secondary	Class	14-17	Intentional	PI	Creation
Benati& Lee (2008)	2008	FL	Japanese	University	Class	>18	Intentional	PI	Manipulation
Benati, Lee & Houghton (2008)	2008	FL	Japanese	Secondary	Class	14-17	Intentional	PI	Manipulation
Benati, Lee & Laval (2008)	2008	FL	Japanese	University	Class	>18	Intentional	PI	Manipulation
Benati (2009)	2009	FL	Spanish	Micellaneous	Class	>18	Intentional	PI	Manipulation
Collentine (1998)	1998	FL	Spanish	University	Lab	>18	Intentional	PI	Creation
De Jong (2005)	2005	FL	Spanish	Micellaneous	Lab	>18	Incidental	N-PI	Manipulation
DeKeyser&Sokalski (1996) - Study1	1996	FL	Spanish	University	Class	>18	Intentional	N-PI	Manipulation
DeKeyser&Sokalski (1996) - Study2	1996	FL	Spanish	University	Class	>18	Intentional	N-PI	Manipulation
Erlam (2003)	2003	FL	French	Secondary	Class	14-17	Intentional	PI	Manipulation
Farley (2001a)	2001	FL	Spanish	University	Class	>18	Intentional	PI	Creation
Farley (2001b)	2001	FL	Spanish	University	Class	>18	Intentional	PI	Creation
Gass& Torres (2005)	2005	FL	Spanish	University	Class	>18	Incidental	N-PI	Creation
Izumi (2002) – Study 1 (without IE)	2002	SL	English	University	Lab	>18	Incidental	N-PI	Creation
Izumi (2002) – Study 2 (with IE)	2002	SL	English	University	Lab	>18	Incidental	N-PI	Creation
Koyanagi (1999)	1999	FL	Japanese	University	Class	>18	Intentional	PI	Creation
Lee & Benati (2007a) - Study1	2007	FL	Japanese	University	Class	>18	Intentional	PI	Manipulation
Lee & Benati (2007a) – Study2	2007	FL	French	University	Class	>18	Intentional	PI	Manipulation
Lee & Benati (2007b)	2007	FL	Japanese	Micellaneous	Class	>18	Intentional	PI	Manipulation
Leeser (2008)	2008	FL	Spanish	University	Class	>18	Intentional	N-PI	Creation
Morgan-Short & Bowden (2006)	2006	FL	Spanish	University	Class	>18	Intentional	PI	Manipulation
Nagata (1998a)	1998	FL	Japanese	University	Computer	>18	Intentional	N-PI	Manipulation
Nagata (1998b)	1998	FL	Japanese	University	Computer	>18	Intentional	N-PI	Manipulation
Qin (2008)	2008	SL	English	Secondary	Class	14-17	Intentional	PI	Creation
Shintani& Ellis (2010)	2010	FL	English	Primary	Class	<13	Incidental	N-PI	Manipulation

Song &Suh (2008)	2008	FL	English	University	Class	>18	Incidental	PI	Manipulation
Takimoto (2009)	2009	FL	English	Micellaneous	Lab	>18	Intentional	PI	Manipulation
Tanaka (2001)	2001	FL	English	University	Class	>18	Intentional	N-PI	Manipulation
Toth (2006)	2006	FL	Spanish	University	Class	>18	Intentional	PI	Creation
Van Patten et al (2009)	2009	FL	Spanish	University	Lab	>18	Intentional	PI	Creation
VanPaten& Wong (2004) - Study1	2004	FL	Spanish	University	Class	>18	Intentional	PI	Manipulation
VanPaten& Wong (2004) - Study2	2004	FL	Spanish	University	Class	>18	Intentional	PI	Manipulation
VanPatten and Cadierno (1993)	1993	FL	Spanish	University	Class	>18	Intentional	PI	Manipulation

## References

- Asher, J. J. (1977). *Learning another language through actions*. Claifornia: Sky Oaks Productions.
- 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(1), 67-93.
- Collentine, J. (1998). Processing instruction and the subjunctive. *Hispania*, 81(3), 576-587.
- DeKeyser, R. M., & Sokalski, K. J. (1996). The differential role of comprehension and production practice. *Language Learning*, 46(4), 613-642.
- Ellis, R. (1997). *Second language acquisition*. Oxford: Oxford University Press.
- Farley, A. P. (2001). Processing instruction and meaning-based output instruction: a comparative study. *Spanish Applied Linguistics*, 5(2), 57-94.
- Gass, S., Mackey, A., & Feldman, L. R. (2005). Task based interactions in classroom and laboratory settings. *Language Learning*, 55(4), 575-611.
- Gass, S., Mackey, A., & Ross-Feldman, L. (2011). Task-based interactions in classroom and laboratory settings. *Language Learning*, 61, 189-220.
- Jeon, E. H., & Kaya, T. (2006). Effects of L2 instruction on interlanguage pragmatic development. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133-164). Amsterdam: John Benjamins.
- Lee, J. F., & Benati, A. (2007). Comparing modes of delivering processing instruction and meaning-based output instruction on Italian and French subjunctive. In J. F. Lee & A. Benati (Eds.), *Delivering processing instruction in classrooms and in virtual contexts: research and practice* (pp. 99-136). London: Equinox Publishing Limited
- Norris, J. M., & Ortega, L. (2006). The value and practice syathesis for language learning. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 3-52). Amsterdam: John Benjamins.
- Qin, J. (2008). The effect of processing instruction and dictogloss tasks on acquisition of the English passive voice. *Language Teaching Research*, 12(1), 61-82.
- Song, M.-J., & Suh, B.-R. (2008). The effects of output task types on noticing and learning of the English past counterfactual conditional. *System*, 36(2), 295-312.
- Takimoto, M. (2009). The effects of input-based tasks on the development of learners' pragmatic proficiency. *Applied Linguistics*, 30(1), 1-25.

- VanPatten, B. (2004). Input processing in second language acquisition. In B. VanPatten (Ed.), *Processing Instruction: Theory, research, and commentary* (pp. 5-32). New Jersey: Laurence Erlbaum Associates
- VanPatten, B., & Cadierno, T. (1993). Input processing and second language acquisition: a role for instruction. *The Modern Language Journal*, 77(1), 45-57.
- VanPatten, B., Inclezan, D., Salazar, H., & Farley, A. P. (2009). Processing instruction and dictogloss: a study on object pronouns and word order in Spanish. *Foreign Language Annals*, 42(3), 557-575.