Effect of explicit and implicit instruction on free written response task performance

Andringa, S.; de Glopper, K.; Hacquebord, H.

DOI
10.1111/j.1467-9922.2010.00623.x

Publication date
2011

Document Version
Final published version

Published in
Language Learning

LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

Effect of Explicit and Implicit Instruction on Free Written Response Task Performance

Sible Andringa
University of Amsterdam

Kees de Glopper
University of Groningen

Hilde Hacquebord
University of Groningen

A classroom study was designed to test the hypothesis that explicit knowledge is used by second-language (L2) learners in a free written response task if that knowledge is present. Eighty-one 12–18-year-old learners of Dutch as an L2 took part in a computer-assisted language learning experiment receiving either explicit or implicit instruction about two grammar structures. The ability to use these structures was measured at three points in time by means of an untimed grammaticality judgment task and a free written response task. Explicit and implicit instruction promoted the use of the target structures in free response tasks equally effectively. However, for one structure, both facilitative and inhibitory effects of explicit instruction were observed if first language similarity was taken into consideration.

Keywords explicit and implicit instruction; form-focused instruction; classroom research; L1 influence; explicit and implicit knowledge; second-language learning

Within the second language acquisition (SLA) literature, form-focused instruction (FFI) research refers to the many studies that have investigated the effectiveness of different types of instruction in different circumstances of learning. R. Ellis (2001) defined FFI as "any planned or incidental instructional activity

We would like to thank Jan Hulstijn and Wander Lowie as well as the three anonymous *Language Learning* reviewers for their meticulous reading of earlier versions of this text and their valuable suggestions for improvement. All remaining issues are our own responsibility.

Correspondence concerning this article should be sent to Sible Andringa, University of Amsterdam, Amsterdam Center for Language and Communication, Spuistraat 134, 1012 VB, Amsterdam. Internet: s.j.andringa@uva.nl

DOI: 10.1111/j.1467-9922.2010.00623.x

that is intended to induce language learners to pay attention to linguistic form." This study focuses specifically on the distinction between explicit instruction (EI) and implicit instruction (II), both examples of FFI. As soon as instruction involves overt attention to the rules of the target language, it must be considered explicit. Conversely, when rules are present but not discussed and learners are not asked to attend to rules during L2 tasks, instruction is considered implicit (Hulstijn, 2005; Norris & Ortega, 2000). This distinction is of particular interest to theories of instructed SLA. Although various labels have been used, these theories generally make a distinction between explicit knowledge (factual and conscious knowledge about rules) and implicit knowledge (procedural knowledge that allows someone to use the second language [L2] appropriately in spontaneous situations of language use) and between explicit learning processes (conscious learning intended to find regularities in language) and implicit learning processes (input processing without the intention to study linguistic form) (Hustijn, 2005). An important issue is how different types of instruction affect the explicit and implicit knowledge systems and learning processes. The goal of this study is to explore how successful EI is in promoting the use of grammatical features in free response tasks in comparison to implicit instruction. In this introduction, we will briefly discuss the reasons why EI might be more effective than II and what research design requirements are needed to demonstrate this. Then we will review the evidence provided by previous research and factors that may affect the results of FFI research.

A consistent finding in FFI research is that an explicit focus on L2 forms promotes their acquisition. This is the picture that emerges from several reviews and meta-analyses of FFI studies (DeKeyser, 2003; R. Ellis, 2007; Larsen-Freeman, 1995; Larsen-Freeman & Long, 1991; Lightbown, 2001; Long, 1983; Norris & Ortega, 2000). Although there has been a bias in the kinds of tests that researchers have used to measure progress in proficiency in favor of explicit kinds of instruction (Norris & Ortega, 2000), a recent review by R. Ellis (2007) demonstrates that EI also positively affects performance on tasks that involve free response or more spontaneous kinds of language production. Roughly, this effect can be explained in two ways. First, EI may trigger concomitant implicit acquisition processes, as suggested by Hulstijn (2002). This would mean that EI works in the same way as II does: It exposes language learners to exemplars on the basis of which implicit acquisition may take place. The second explanation would be that the explicit knowledge resulting from EI somehow provides added value to the SLA process. This argument has been advanced by several researchers in different forms within the framework of the interface debate (e.g., DeKeyser, 1998; N. C. Ellis, 2005; N. C. Ellis & Larsen-Freeman, 2006; R. Ellis,

1994; Krashen, 1981, 1985, 1994; Paradis, 1994, 2004). The two explanations are probably not mutually exclusive: Positive effects of explicit instruction on free response tasks may be due to a combination of both concomitant implicit acquisition processes and positive contributions of explicit knowledge.

The discussion about the pedagogical value of EI and explicit knowledge was sparked by Krashen (1981, 1985, 1994). He maintained that explicit and implicit knowledge are separate, noninterfacing systems, that L2 proficiency is based on implicit knowledge only, and that EI contributes little to the development of L2 proficiency because it leads to explicit learning and knowledge rather than implicit acquisition. Opposing views have been put forward by, among others, DeKeyser (1998), Hulstijn (1999), O'Malley, Chamot, and Walker (1987), Schmidt (1990, 1995), and Sharwood-Smith (1988), who claimed that explicit knowledge is available for use in linguistic tasks and can be proceduralized, and Bialystok (1989, 1994), who argued that language learning entails a process of developing awareness for linguistic structure. In these views, EI and explicit knowledge may be necessary in certain circumstances. Additionally, there are those who see explicit and implicit knowledge and learning as separate but argue that EI may help learners to notice gaps in their interlanguage, thus facilitating implicit acquisition (Doughty & Williams, 1998; N. C. Ellis, 2005; R. Ellis, 1997; Long & Robinson, 1998; Skehan, 1998). In this view, teachers should provide EI in the form of brief shifts away from meaningful interaction when they perceive difficulties. Finally, a special position is taken by Paradis (1994, 2004). On the basis of neuroimaging studies, he claimed that L2 proficiency is based on proceduralized explicit knowledge rather than implicit knowledge.

The role of FFI studies in resolving this discussion is limited. This would require pure measures of explicit and implicit knowledge as well as measures that provide insight into the processes triggered by a particular kind of instruction. The outcome measures used in FFI cannot reveal how knowledge systems interacted and what processes were triggered. However, if FFI studies demonstrated that EI is more effective in improving free response task performance in comparison to II, then that would imply that EI and II lead, at least partly, to different kinds of knowledge and that there is added value to EI and explicit knowledge of the rules of the target language. The strongest evidence would be provided by studies that compare performance by groups with and without explicit knowledge of a particular target structure while controlling for the amount of exposure. This is where FFI research tends to fall short. In many studies investigating the effect of EI, explanation of rules constituted an extra activity or it was matched by an activity that did not necessarily provide

the same amount of intensified exposure to the target structures (e.g., Day & Shapson, 2001; Mackey & Philp, 1998; Muranoi, 2000; VanPatten & Sanz, 1995). As a result, the superiority of a particular type of instruction may often have been confounded by exposure differences, which are not often controlled for in FFI studies. EI inevitably provides exemplars to the L2 learners that may trigger implicit learning processes. For this reason, one should contrast an EI group with an II condition exposing language learners to the target structures in equal amounts.

Another research design requirement is that studies should use both free response measures and measures that draw upon explicit, declarative knowledge. This, too, is challenging because there are no accepted measures of explicit and implicit knowledge. In an attempt to define explicit and implicit knowledge, R. Ellis (2004) separated the two on several dimensions, defining implicit knowledge as essentially intuitive, procedural, systematic, automatically accessible in situations requiring fluent use, and nonverbalizable and defining explicit knowledge as conscious, declarative, potentially incorrect and incomplete, consciously accessible—especially when problems arise in decoding or producing the linguistic form (e.g., because an incorrect form is encountered or when an utterance is extremely complex)—and verbalizable. Based on these characterizations, Ellis considered tasks requiring the use of the target structure under the constraints of natural language use, such as free response tasks, tests of implicit knowledge. Tasks that involve some kind of metalinguistic judgment, that are not time-pressured, that present target structures in isolated contexts, and in which the structure is presented incorrectly are most likely to tap into explicit knowledge, according to Ellis. There is some empirical evidence that these two kinds of knowledge can be separately tested. In studies by Han and Ellis (1998) and R. Ellis (2005), a number of timed and untimed tests of grammar were administered. The oral narrative task used in their study was found to load on one factor together with the time-pressured grammaticality judgment task, whereas the metalinguistic knowledge test made up a second factor with the untimed grammaticality judgments. These results suggest that explicit and implicit knowledge are, to some extent, separate systems that can be separately tested.

Claims of the superiority of EI over II are mainly based on studies using measures of controlled production. Recent reviews (DeKeyser, 2003; R. Ellis, 2002, 2007) and meta-analyses (Norris & Ortega, 2000; Spada & Tomita, 2008, 2010) testify to the fact that relatively few FFI studies have made use of free response measures to assess the effects of EI; even fewer studies have used these measures and compared explicit and implicit types of instruction.

In DeKeyser's (2003) review, 14 studies were included comparing explicit and implicit treatments in both laboratory and classroom studies; he concluded that EI was much more effective than II, but he also pointed out that none of the studies used free response measures: "the dependent variable has always been a test that allows for some degree of monitoring of explicit knowledge" (p. 326). The meta-analyses by Norris and Ortega (2000), and more recently Spada and Tomita (2008, 2010), have also been cited in support of claims of the superiority of EI over II. However, in Norris and Ortega, for example, 10 studies compared explicit and implicit types of instruction; of these, only 1 made use of a free response measure. In addition, the effect sizes reported by the meta-analyses are based on very different tests even within the defined categories, using different measurement scales, and assessing mastery of many different properties of language. Additionally, for the purpose of meta-analysis, the number of studies included was quite small (Chaudron, 2006; N. C. Ellis, 2006). All in all, there is convincing evidence that EI is generally superior to II when measures of controlled production are used. However, for measures of free production, the evidence is circumstantial at best.

Of the FFI studies included in the reviews and meta-analyses of Norris and Ortega (2000), DeKeyser (2003), and Spada and Tomita (2008), three studies compared explicit and implicit treatments and used free response measures: These are Sanz and Morgan-Short (2004), Muranoi (2000), and Williams and Evans (1998).[1] All three will be examined more closely here. The study of Sanz and Morgan-Short, investigating the role of EI in SLA, meets the criteria necessary for demonstrating the superiority of EI over II best. They compared explicit with implicit conditions and used free as well as more constrained measures of progress. In this study, the instruction was delivered by computer and involved implicit practice in Spanish preverbal pronouns. The participants were university students learning Spanish in the United States. Using a between-groups design, Sanz and Morgan-Short compared four different treatment conditions. In all conditions, the students were faced with sentences containing the target structure and were asked to respond to them appropriately by clicking the picture corresponding with the sentence or by choosing the correct translation. A correct response depended on correctly processing the target structure. In one condition, practice was accompanied by explicit rule explanation; another condition involved explicit feedback during the practice; the third condition featured both rule explanation and feedback; and the fourth featured none of these. Each group consisted of approximately 17 participants. Progress was measured by means of a sentence completion task and a written video retelling task. In the sentence completion task, participants were asked to use a particular

verb to finish the sentence appropriately; the task was not time-pressured. Sanz and Morgan-Short found all four groups to improve significantly on both the sentence completion task and the video retelling task, but there were no interactions between the type of instruction received and the progress students made. One interpretation of this result could be that EI is not superior to II in free response tasks. Alternatively, the EI may have been insufficiently effective to have constituted an advantage. If the sentence completion task in this study is taken to be a measure of explicit knowledge, then the results indicate there were no differences in explicit knowledge between the groups compared: A gain was demonstrated in both conditions. As a result, positive effects on the free response measure because of differences in explicit knowledge cannot be expected.

The study by Muranoi (2000) was designed to investigate the effect of interaction enhancements. Using 90 Japanese university students learning English, Muranoi compared three types of instruction: interaction enhancement with form-focused debriefing (IEF), interaction enhancement with meaning-focused debriefing (IEM), and no interaction enhancement with meaning-focused debriefing (NEI). Interaction enhancement involved providing students with scenarios that were intended to create contexts for using indefinite articles. During role-play, the teacher would enhance the interaction by providing implicit negative feedback (i.e., repetition requests and output modification). During the form-focused debriefing, the students received EI in how to use the indefinite article, whereas the meaning-focused debriefing discussed the success of the interaction. Progress was assessed by means of oral story retelling, oral and written picture description, and a grammaticality judgment task. The results showed that the IEF group scored significantly higher on all four measures than the IEM and NEI groups. In other words, the group with the most explicit knowledge also showed most progress on the free response measures. Muranoi's results are a first indication that having explicit knowledge constitutes an advantage in performance on free response tasks. However, it is important to point out that the IEF group was exposed more to the target structure, because of the differences in debriefing. As a result, the advantage found for the IEF group cannot safely be attributed to the fact that IEF had more explicit knowledge.

Another study comparing the effects of explicit and implicit instruction was conducted by Williams and Evans (1998). It addressed the issue of whether the kind of grammar structure makes a difference in instruction. Their learners of English as a second language (university students) were divided into three conditions: one that received a so-called input flood—frequent exposure to the target structure without referring to it—operationalized in Williams and

Evans's study as written input that was manipulated to contain many examples of the target structures. Another group received an input flood plus explicit rule presentation; and there was a control group that did not receive any instruction. Each group consisted of 11 participants. The instruction targeted participial adjectives of emotive verbs and passive constructions. Sentence completion and narratives were used to measure progress. For this review, only the passives are of interest, because the narrative task was used for this structure only. It should also be noted that the narrative task was not a true free response task because the subjects were asked to describe pictures and were supplied with a particular phrase eliciting the passive to start off the narrative. Results showed that the control group hardly made progress. Both instructed groups did make progress, but there were no significant differences between the two on the narrative task. The flood plus instruction group seemed to make more progress on the sentence completion task, but significance was not reached. Both experimental groups outperformed the control group on the narrative task, suggesting that intensified exposure to the target structure leads to better performance. As such, the study does not provide evidence for advantages in free response task performance due to EI and explicit knowledge. There are several explanations for these findings. Perhaps there were no differences in explicit knowledge between the groups compared, which would mean that either the II lead to explicit knowledge or the EI failed to result in explicit knowledge. It may also be that the explicit group for some reason did not have the opportunity to take advantage of their knowledge in the narrative task.

So far, the effects of EI and II have been considered irrespective of potentially intervening factors. What complicates FFI research are the many factors that may strengthen or attenuate the effects of instruction. Several studies have investigated how the nature of the target structure may affect success of instruction (e.g., De Graaff, 1997; Robinson, 1996). Most have looked at what DeKeyser (2003) referred to as "objective difficulty": features inherent to the structure that make it more or less complex. Examples of such features are functional and formal complexity (DeKeyser, 1998, 2005), reliability and scope (Hulstijn & de Graaff, 1994), the linguistic nature of the target structure (morphological, syntactic, or formulaic; R. Ellis, 2002), and features related to frequency and salience (Goldschneider & DeKeyser, 2001). The findings sometimes seem contradictory. DeKeyser (1995) contrasted categorical and probabilistic rules, finding that the categorical rules were learned significantly better through EI, whereas the probabilistic rules were promoted nonsignificantly better by II. Robinson (1996) investigated structure complexity and found EI to be more effective for the simple rule and no differences for the

complex rule. De Graaff (1997) also looked at complexity, but found explicit instruction to be more effective for complex rules. The aforementioned studies used measures of controlled production. R. Ellis (2002) reviewed the impact of FFI instruction on free response measures and concluded that instruction was mainly effective when the targeted rules were morphological or formulaic in nature. It should be noted that the instruction was mostly implicit in the studies reviewed by Ellis. Interpreting these findings is difficult. The most powerful interpretation was probably provided by DeKeyser (2003). On the basis of a review of studies that investigated the relation between explicit and implicit learning, DeKeyser suggested that structures that are difficult to learn by means of simple association, such as arbitrary form-meaning connections, cannot be learned easily implicitly and will require more explicit learning processes. This would mean that more abstract and arbitrary structures would benefit from more explicit kinds of learning and instruction, whereas structures that are concrete and observable in the input can be taught implicitly effectively. This interpretation seems to converge rather well with findings so far.

Not only might a structure's objective difficulty affect the success of instruction, its subjective difficulty may do so as well. Subjective difficulty refers to the notion that L2 learners may differ in how complex they perceive a structure to be (DeKeyser, 2003). Differences in aptitude and first language (L1) background, for example, may affect a person's ability to acquire a particular L2. Aptitude and L1 background have been identified as important on the basis of theories of acquisition or reviews of research, but they have mostly not been systematically investigated. Robinson (2002) has suggested that aptitude is a multicomponential construct and that some components index the ability to learn explicitly, whereas others are related to the ability to learn incidentally or via focus on form. The claim that the L2 learner's L1 affects the perception of instruction has been made by, among others, Doughty and Williams (1998), Harley (1989), and VanPatten (1996, 2003). VanPatten argued that L1 processing strategies may put L2 learners on the wrong track; instruction would be effective when it informs L2 learners about mismatches between the L1 and L2. Processing Instruction, as it is called, has been implemented quite frequently, and positive effects have been observed in comparison to various control conditions (Salaberry, 1997; Sanz & Morgan-Short, 2004; VanPatten & Sanz, 1995).

Having reviewed the empirical literature, one must conclude that FFI research has not provided much evidence that EI and explicit knowledge offer added value to L2 learners in free response tasks. Although there is ample evidence that EI leads to improved performance on such tasks, it is still unclear

whether EI is more effective than implicit treatments, which would support the idea that explicit knowledge is used in free response tasks. To be more precise, in the few studies that have compared EI and II in combination with a free response task, EI was found to be effective, but none of the studies reviewed allow for the interpretation that EI was superior to II. Either there were no differences between the explicit and implicit conditions on the free response tasks (e.g., Sanz & Morgan-Short, 2004; Williams & Evans, 1998), or there were substantial differences between the conditions in exposure to the target structure (e.g., Muranoi, 2000), or the results were difficult to interpret because there were no differences on the task that might indicate differences in explicit knowledge (e.g., Sanz & Morgan-Short, 2004). The latter could mean that the task used was not a good test of explicit knowledge (it was not intended as such); it could also mean that the explicit instruction was ineffective or that the II lead to explicit knowledge. The present study explored whether EI about two structures promoted their correct use in a free written response task more effectively than II, whether the success of EI depended on the nature of the target structure, and whether the L1 background affected the success of instruction.

Method

Design

In the present classroom study, learners of Dutch as a second language took part in a computer-assisted language learning experiment in which they received either EI or II about two grammar structures. To avoid bias due to exposure differences, the amount of input and the linguistic context in which the input was offered was precisely matched, the only difference being the degree of explicitness in presentation of the target structures. Progress was measured at three points in time (immediately before the instruction [T0], immediately after [T1], and with a delay of 4 weeks [T2]), each time by means of an untimed grammaticality judgment task and a free written response task. Two target structures were chosen—the degrees of comparison (DoCs) and verb-final in subordinate clauses (V-final)—to test the hypothesis that the effectiveness of instruction depends on the nature of the grammar structure. The study adopted a crossed-treatment design: participants in treatment group 1 received explicit instruction about the DoCs and implicit instruction about V-final; for treatment group 2 this was the other way around. Questionnaires and additional tests were administered to assess motivation, aptitude, style, age, and educational experience. The participants' L1s were coded for how these languages expressed the meanings of the structures under investigation. This design allowed for an

investigation of the effects of EI and II, whether such effects are different for contrasting target structures, and whether effects of instruction depend on perceived difficulty of the target structures due to L1 differences. The following hypotheses were tested:

1. Explicit instruction about the degrees of comparison leads to better performance on a free written response task as compared to II.
2. Explicit instruction about verb-final in subordinate clauses leads to better performance on a free written response task as compared to II.
3. The success of EI and IIdepends on the way the target structures' meanings are expressed in the participants' L1s.

Participants

The subject sample for this study consisted of 102 students of Dutch as a second language from eight secondary schools in the north and west of the Netherlands. These schools have special facilities to train immigrant students ranging between 12 and 18 years of age, providing intensive Dutch courses to enable their students to follow regular secondary school programs as soon as possible. The courses consisted of about 4–5 hr a day of Dutch training. Grammar is not very prominent in these courses because they almost all used a course book that does not introduce grammar explicitly until well into the second year of training. Students did not follow additional courses or lessons, but there may have been variations between schools and teachers. The schools were selected based on proximity, availability of students, and willingness to cooperate. The selection of participants for inclusion in the study was based on their overall L2 proficiency. The so-called ISK tests (Schuurs, 1999) formed the most important tool for participant selection because most schools used this test as a placement test and to monitor progress in Dutch. They are a series of standardized tests of listening, speaking, writing, and reading, developed especially for assessing Dutch language proficiency of adolescent learners of Dutch as a second language. Participants scoring approximately 50 points (out of 100) on this test were considered fit for participation: 50 points corresponds to approximately 1 year of ISK training and a sufficient level of Dutch proficiency to participate in this study. ISK test scores were obtained for 88 participants. Their mean score was 53.1 ($SD = 5.2$), and observations ranged between 41 and 67. In total, 102 students were tested at time 0 (T0); 81 students finished the instruction and made the tests at time 1 (T1); 67 students were available at time 2 (T2). Most analyses in this study were performed on complete data, reducing the number of participants effectively to 67.

The participating students were immigrants from many different countries who had come to the Netherlands for various reasons. The sample included speakers of 33 different languages, the most frequently occurring language being Portuguese as spoken in Angola ($n = 23$). There were no other languages with more than 10 speakers present in the subject pool. The only Germanic language present was English, spoken by 5 participants; 11 participants were from Europe. To check for L1 influence, specific features of the participants' L1s were coded that might affect their use of the target structures. For the degrees of comparison, the participants' L1s were coded for how they express comparison: morphologically or periphrastically. A chi-square analysis revealed no significant differences in the distribution of morphological and periphrastic L1s between the students in the EI and II conditions. For subordination, the participants' L1s were checked for the presence of similar inversion in subordinate clauses. None of the L1s present in the sample had this feature. The L1 was coded for basic word order (SVO or SOV) on the basis of the hypothesis that participants with SOV in their L1s might find it easier to apply SOV in Dutch subordinate clauses. A chi-square analysis on the distribution of SVO and SOV languages did not reveal significant differences between the two conditions.

The participants were asked if they knew any L2s. Of the 81 students included in the study, 36 reported not knowing any L2s. If participants indicated knowing an L2, they mostly mentioned Arabic, French, Spanish, or a regional language. In total, 11 students reported knowing English as a second language, and 1 reported German. In the case of English and German, for which the degrees of comparison are similar, all indicated that their level of Dutch as a second language was higher than their level of English or German.

At T0, a number of additional tests were administered. These were a C-test, a rote memory test for verbal material, and a grammatical sensitivity test that were developed for the occasion. The memory test and the grammatical sensitivity test were based on components of the Modern Language Aptitude Test (Carroll & Sapon, 1959). The procedures were the same, but Finnish words were used for the memory test (the MLAT made use of Kurdish; we expected that Kurds might participate in this study) and for the grammatical sensitivity test, a Dutch version was developed. In addition, teachers were asked to fill in questionnaires about their students' age, amount of exposure to the L2, length of stay in the Netherlands, length of stay at school, cognitive style, motivation, and aptitude. At T0, significant differences were not observed among the groups for any of these variables. This was also true for the subsample at T1 (81 students) and T2 (67 students), indicating that the loss of students was either random or similar in both groups.

Table 1 Examples of the degrees of comparison in Dutch

Funny:	grappig	grappiger	grappigst
Nice:	leuk	leuker	leukst
Sweet, dear:	lief	liever	liefst

Target Structures

A small-scale, intensive pilot study was conducted with 20 students who did not participate in the main experiment to find out which structures were emerging in the interlanguage of the students. The pilot required the students to write short pieces of text and was intended to evaluate how well the target structures could be elicited implicitly. The degrees of comparison (DoCs) and verb-final in subordinate clauses (V-final) were chosen as target structures for the main study because they were found to emerge more or less simultaneously in the students' interlanguage and because they could be elicited with considerable success.

The Dutch DoCs are very similar to the English degrees of comparison. The comparative and superlative are formed by attaching *–er* and *–st* to the adjective or adverb (see Table 1). Both suffixes carry independent meaning and modify the meaning of the adjective or adverb to which they are attached. Because each suffix carries a straightforward meaning, the constructions must be considered functionally simple. However, Dutch also allows for periphrastic DoCs by means of the use of the quantifiers *meer* (more) and *meest* (most). Periphrastic use is restricted to a small number of adjectives that can be used predicatively only, to complex adjectives for which pronunciation difficulties would arise (*meer gebruikelijk*: more usual), and to some present and past participles (*meest verdorven*: most perverse); in the vast majority of cases, using periphrastic DoCs is ungrammatical. Periphrastic comparison was not subject of the instruction, simply because the kinds of adjectives that are realized periphrastically in Dutch are generally not part of the vocabulary of the L2 learners participating in this study. Structures may be difficult for L2 learners because of their meaning or their form or because it may be difficult to map meanings onto forms (DeKeyser, 2005). Difficulties L2 learners might have in acquiring the Dutch DoCs would probably not be caused by the structure's meaning: All languages appear to possess ways to express DoCs. However, the structure's form may be difficult: Morphological rules may be difficult for L2 learners simply because they are easily overlooked (DeKeyser, 2005). Finally, mapping meaning onto the forms (*–er* and *–st*) may also be difficult for learners. The forms are never optional; however, they are sometimes redundant or opaque.

The opacity of the forms may be affected by the fact that adjectives are also marked for gender (*de grootste* "the biggest," *het kleinst* "the smallest"). In addition, Dutch also uses the *–er* suffix to make nouns of verbs (*werken* "to work," *werker* "worker"), which also reduces the form's opacity. Both the *–er* and *–st* suffix may be redundant. The *–er* suffix is redundant in phrases including a comparative clause (*Ik ben langer dan Anne* "I am taller than Anne"); superlative forms in predicative position are always accompanied by definite articles (*het leukst, de grootste*). Additionally, suffixation sometimes causes changes in the spelling of the adjective or adverb (see the third example in Table 1). These changes are in accordance with Dutch spelling conventions. The structure must probably be considered large in scope—because there are many adjectives that can be inflected—and reliable—because periphrastic comparison is not very frequent in Dutch and occurs almost exclusively with adjectives that probably lie outside the L2 learners' vocabulary.

Subordinate clauses are introduced by subordinating conjunctions that express how the main clause and the subordinate clause relate to each other. Subordinating conjunctions can express, among others, temporal, causal, and conditional relations. Particularly difficult for L2 learners is that the finite verb—while located before the object in main clauses (VO)—is positioned after the object (OV) in subordinate clauses (see Table 2). Although this repositioning of the finite verb is obligatory, neglecting to reposition the verb does not lead to incomprehension or miscomprehension. The problem that learners are faced with is primarily a problem of form. L2 learners might try to map meaning onto the form by associating the rule with the relationship the clause intends to express, and in doing so, they will find the rule redundant because all meaning is expressed by the subordinating conjunction. Things are

Table 2 Examples of subordinate clauses in Dutch (taken from the data collected for this study)

1.	Main clause:	Mijn vriend vertelt een grap.
	Subordinate clause:	. . . , als mijn vriend een grap vertelt.
		. . . , *when my friend a joke tells.*
2.	Main clause:	Zij wil een ijsje eten.
	Subordinate clause:	. . . , omdat zij een ijsje wil eten.
		. . . , *because she ice-cream wants to eat.*
3.	Subordinate clause:	. . . , omdat de kleine radio goedkoper is.
	Coordinate clause:	. . . , want de kleine radio is goedkoper.
		. . . , *because the little radio (is) cheaper (is).*

further complicated by the fact that subordinate conjunctions sometimes have coordinate counterparts expressing the same meaning. For example, causality can also be expressed by means of the coordinating conjunction: *want*. In Table 2, example 3 illustrates two sentences that are synonymous in meaning but not in word order. The presence of a subordinate conjunction is the only valid and strong cue for V-final. The occurrence of one of the subordinating conjunctions leads to verb repositioning almost without exception. The only exception may be when the subordinate clause is too complex to move the verb all the way to the end—for example, when the subordinate clause is very long or because there is a subordinate clause within the subordinate clause (*Hij zegt dat Anna weet dat hij ziek is* "He says Anna knows he is ill"). At the participants' proficiency level, this would not be likely to occur. The applicability of the rule is virtually endless, and especially conditional and causal subordinate clauses occur rather frequently in everyday language.

Instruction

The goal of the instruction was to create between-group differences in explicit knowledge while controlling for the amount of exposure to the target structures. To this end, EI and II conditions were created for both structures. The two conditions were largely in accordance with Housen and Pierrard's (2005) characterization of explicit and implicit FFI. According to this characterization, implicit FFI attracts rather than directs attention, is minimally obtrusive, occurs spontaneously, presents the structure in context, does not involve metalinguistic terminology, and encourages free use. EI directs attention to the target structures, is a planned activity, interrupts communication, discusses structures in isolation, makes use of metalinguistic terminology, and involves practice (Housen & Pierrard, 2005). The instruction was delivered by means of a computer program that was developed for the occasion in Authorware (1999). A series of eight lessons was developed that focused on three different themes: Advertising, the Olympic Games, and Water in the Netherlands. Each lesson consisted of short texts about these themes, and students also received some unfocused exercises related to the texts. These texts and exercises would be interlaced with focused exercises, offering form-focused instruction in the DoC or V-final either implicitly or explicitly. As there were eight lessons in total, four focused on the DoC and four focused on V-final. Whether this focus was explicit or implicit depended on the treatment group the students were in. The input in both conditions was precisely matched, meaning that the same input sentences were used to create either an explicit or implicit focus on the target structures. In practice, input matching was implemented by modifying

text comprehension exercises into grammar exercises. In this way, the original implicit focus on the target structures was changed into an explicit focus and the input itself remained unchanged. Thus, the two treatment groups were exposed to virtually the same input, meaning that they had more or less equal chances to learn implicitly from the instruction. Table 3 exemplifies this design. More information about the instruction can be found in Andringa (2005).

Despite the efforts to keep the input between the conditions similar, there were some differences. One obvious difference involves the rule explanation that was part of the EI. The aim of the EI was to create an explicit understanding about the target structures. EI was accompanied by examples and, if possible, visual aids. Care was taken not to make the explanation too technical by limiting the use of specific linguistic terminology, and when used, such terms were carefully defined. Although the use of the computer limited the possibilities, several kinds of exercises were used to strengthen the students' explicit knowledge. In the receptive phase of the instruction, students had to recognize the target structure in example sentences by means of multiple choice or yes/no questions. Input enhancement was also used in dialogues in which the target structure was underlined and explicitly referred to. As the students progressed, more productive exercises were used, such as gap-filling and sentence completion. Explicit exercises were clearly marked by a special logo, caption, and background color. In addition, the exercises were introduced by one or two screens of rule explanation, either introducing new aspects of the rule or repeating aspects presented earlier. In the subsequent exercises, the participants were continuously reminded to apply the rules just explained.

The II was created to invite the students to process the target structure for meaning. However, the students were never made aware of this; there were no explicit attempts to get the language learner to notice the target structure that would be abundantly present in the comprehension exercises. Providing the proper answer thus forced the participants to process the target structure for meaning. These exercises were always based on one of the texts just read and required remembering or checking the facts in the texts. In later stages of the program, the participants were also asked to type their answers to questions that were designed to elicit the target structures. To compensate for the extra exposure to the target structures in the explicit condition, students in the implicit condition received extra input during the exercises themselves—for example, by means of feedback screens or extra exercises.

The instruction was relatively short. On average, participants spent 3–4 hr on the program, mostly in three or four 1-hr weekly sessions. Their teacher organized the computer sessions and oversaw that all participants finished the

Table 3 Examples of input matching (translated from Dutch)

Explicit instruction	Implicit instruction
Degrees of comparison (DoCs)	
Which word is a form of the DoCs?	This statement is about the text you just read. Is it true or false?
When a commercial is funny, people think it is nicer.	When a commercial is funny, people think it is nicer.
Degrees of comparison	
These three sentences are taken from advertisements. Which sentence does not contain a form of the degrees of comparison?	What does this advertisement try to tell you?
A – X now washes even cleaner.	A – X now washes even cleaner.
B – With X, colors stay nicer.	B – With X, colors stay nicer.
C – X is not expensive.	C – X is not expensive.
Degrees of comparison	
Fill in one of the (three) words in the sentence below. Use a form of the DoCs:	Fill in one of the (three) words in the sentence below. Make the sentence agree with the text:
X is one of Holland's [high/fast/small] skaters.	X is one of Holland's [high/fast/slow] skaters.

(*Continued*)

Table 3 Continued

Explicit instruction	Implicit instruction
Subordinate clauses (V-final) Which sentence does not have a subordinate clause? A – X likes the Sony, because it is very small. B – X likes the Sony, because it looks smart. C – X likes the Sony: it is very small and it looks smart.	Why does X like the Sony discman? A – X likes the Sony, because it is very small. B – X likes the Sony, because it looks smart. C – X likes the Sony, because it is very small and it looks smart.
Subordinate clauses (V-final) Make from two sentences one complex sentence: 1 - You can earn a lot of money. 2 - You are good in sports. You can earn a lot of money if . . .	Read the text (excerpt given) and finish the sentence: You can earn a lot of money if . . .

program. Teachers were instructed to offer minimal help and not to discuss the forms in class over the course of the experiment. Any association between the tests and the instruction was avoided as much as possible.

Measures

The students' ability to use the target structures was assessed with two tests: a grammaticality judgment task (GJT) and a free written response task (FWRT). Both tests were paper-and-pencil tasks and were identical at each time of measurement. They were administered by one of the researchers.

The purpose of the GJT was to check for differences in explicit knowledge between the two conditions. Several reasons motivated the choice of a GJT. First, as already pointed out, research by Han and Ellis (1998) and R. Ellis (2005) suggested that this task taps into explicit knowledge, especially those items that contain errors and provided that there is no time pressure. Another reason to choose a GJT was that the task can call on explicit knowledge without making explicit reference to the target structures in focus, which would be advantageous to students who had had EI and which would have made students in the II condition aware of our focus. In the GJT that was used for this study, the participants were given target sentences containing incorrect forms of the target structure and were instructed to indicate whether they recognized any errors in the sentence. In addition, they were asked to underline the error they had spotted. A response would be scored correct only if the error was properly underlined in the sentence. A sentence could contain only one error. For the DoCs, subjects had to see that a particular adjective was not marked for comparison while the context required this. For subordinate clauses, V-final was violated. Examples are provided in Table 4. The test consisted of 20 items in total. Of these, only one item was correct. A small-scale pilot (18 participants) with a 40-item GJT had revealed practical problems. Because the subjects of this study were selected to have little knowledge of the target structures, taking this test proved simply too demanding and too time-consuming. The test contained more correct items but frustrated the participants: They got the feeling that they had missed many errors. In the light of this experience, it was decided

Table 4 Grammaticality judgment task example (translated from Dutch)

1. Anne is lang, maar Lien is nog lang.	correct/incorrect
Anne is quite long, but Lien is even long.	
2. Ik ga naar huis, omdat ik heb veel te doen.	correct/incorrect
I am going home, because I have a lot to do.	

Table 5 Reliability coefficients (KR-20) of the grammaticality judgment test

	T0 ($n = 102$)	T1 ($n = 81$)	T2 ($n = 67$)
Degrees of comparison	.70	.80	.81
Verb-final	.78	.86	.78

to keep the number of test items small and to have only one correct sentence. The incorrect items consisted of six sentences with incorrect instances of the DoCs and eight sentences with incorrect instances of subordinate clauses. Five sentences contained errors unrelated to the target structures, and one sentence was correct. The errors in the sentences not focusing on one of the target structures were kept as simple as possible. Separate GJT scores were calculated for the DoCs and V-final. Table 5 provides the reliability coefficients for both the DoCs and subordinate clauses. These were rather good, indicating that the tests were internally consistent.

The purpose of the FWRT was to evaluate the students' ability to use the target structures incidentally and under the constraints of normal language production. A written production test was chosen for several reasons. Most importantly, the effects of instruction are expected to affect written proficiency before they do oral proficiency. Being less time-constrained, L2 learners have more opportunities to apply explicit knowledge in their output. Studies by Day and Shapson (2001) and VanPatten and Sanz (1995) confirmed this; they found no effect of their instruction on oral proficiency tests, but instruction did affect written proficiency. Thus, a written test probably maximizes chances of finding an instruction effect. A more practical consideration was that collecting and processing written data is simply faster and less labor-intensive, thus allowing for a larger subject sample. The task required students to read a short text describing a particular situation or to look at a particular comic strip or picture. Questions were asked about these, and the students had to respond appropriately in one or two sentences. There were no elicitation phrases or an explicit invitation to use the target structures. The test was presented as a writing proficiency test; students were asked to write carefully and correctly.

Test items that intended to elicit forms of the degrees of comparison placed the test-takers in a situation in which they had to make comparisons. For example, they were asked to compare two radios or to choose between two pairs of jeans. No attempt was made to specifically elicit the use of either comparative or superlative forms. Elicitation of subordinate clauses aimed primarily

for conditional and causal subordinate clauses, being by far the most frequent in Dutch. In addition, the pilot study had revealed that subordinate clauses expressing relations other then conditional or causal relations were very difficult to elicit. Conditional subordinate clauses can be elicited effectively by means of questions asking when (under what condition) something may occur. For example, questions starting with the Dutch question word *wanneer* (meaning *when* or *under what circumstances*) resulted in high rates of conditional subordinate clause use. Situations aiming for the use of causal subordinate clauses proved to be less compelling. Students were asked to explain what caused a particular thing to happen or why they would make a particular choice. For example, when asked why they would choose a particular radio, students might answer "because it is cheaper." However, in the pilot, participants also frequently omitted the subordinate conjunction "it is cheaper" or avoided the use of the subordinate conjunction by means of the synonymous coordinate conjunction *want*: "[want] it is cheaper" (see also Table 2).

In total, the test consisted of 29 items. Eight items focused uniquely on the DoCs; 11 elicited V-final; and 10 items invited the use of both structures. Consequently, there were 18 opportunities to use the DoCs and 21 opportunities to use V-final. All items were selected on the basis of a pilot study with 11 highly proficient L2 learners. The items were considered fit for inclusion if at least eight participants had used the target structure. The FWRT scores simply consisted of the number of correct occurrences of the target structures. Each occurrence of the target structure was included; that is, if an item generated multiple uses of a target structure, each use would be counted. Errors in spelling were disregarded. For the purpose of calculating test reliability coefficients, a item was scored *correct* if it elicited at least one correct instance of the target structure; incorrect use, no use, or avoided use in obligatory contexts was scored as *incorrect*. Table 6 provides the reliabilities for both target structures.

The GJT and the FWRT scores were significantly correlated. For the DoCs, correlations of $r(100) = .54$, $r(79) = .55$, and $r(65) = .54$ were found at respectively T0, T1, and T2. For V-final, these correlations were $r(100) = .50$,

Table 6 Reliability coefficients (KR-20) of the free written response task

	T0 ($n = 102$)	T1 ($n = 81$)	T2 ($n = 67$)
Degrees of comparison	.83	.83	.77
Verb-final	.91	.90	.89

$r(79) = .52$, and $r(65) = .52$. Apparently, to some extent, the tests measured the same construct.

Procedures

All data were collected over a period of 3–4 months. Once the program was installed and the students had been pretested, the instruction started. The pretest (T0) was not always immediately followed by the instruction. In some schools, technical problems led to a delay of 2 or 3 weeks before the students started with the instruction. The students worked individually on separate PCs in computer classrooms, supervised by their regular teacher. To start the program, students had to enter their login number and were assigned randomly by the computer to one of the two treatment groups. The students were retested about 4 weeks later when they had finished the program (T1) and again 4 weeks after that (T2). All tests were administered in class by the first author, mostly in the absence of the students' teacher. The students were told that they participated in a study into the acquisition of Dutch as a second language and that their Dutch would be monitored for this purpose. In addition, they were reassured that the tests did not serve any additional purposes related to their school career. No mention was made of grammar or the specific grammar structures in focus. Nor was the computer program mentioned that was part of the research. The FWRT was always given before the GJT: Although there were no explicit references to the target structures in the GJT, the task might call the participants' attention to the structures.

Univariate analysis of covariance and repeated-measures analysis of variance were used to test our hypotheses.[2] Two dependent variables were used: the GJT scores and the FWRT scores, both obtained at three different points in time. The factors Instruction (EI, II) and L1 Similarity (Different or Similar) were operationalized as two-level between-subjects factors. For the repeated-measures analyses, the factor Time was included as a three-level within-subjects factor (T0, T1, and T2). All analyses were performed separately for each target structure and each dependent variable. Univariate analysis of variance with repeated-measures requires that the assumption of sphericity is met. In case of violations of this assumption, the most conservative correction was reported: the Greenhouse-Geisser estimate. Analyzing skewness and kurtosis values for each measure demonstrated that the data were not always normally distributed. In fact, at T0, all distributions were positively skewed. Skewness diminished at T1; and at T2, there were no significantly skewed distributions. Because the participants were selected to have little knowledge of the target structures, a positively skewed distribution at T0 is hardly surprising. Violations of kurtosis

occurred as well, but not very often. Although these violations were unfortunate, analysis of variance is robust against such violations (Van den Bercken & Voeten, 2002), especially if these violations are not too grave, as was the case here. Power analyses revealed that a power of 1.0 was reached for the 2 × 3 repeated-measures analyses to detect medium within-subjects effects and medium interaction effects ($f = .25$) in terms of the standards set by Cohen (1988).

Results

Results for the Degrees of Comparison

The following analyses pertain to the first hypothesis: EI about the degrees of comparison leads to better performance on the FWRT as compared to II. Mean scores and standard deviations on the GJT and the FWRT were computed for the three times of measurement, as shown in Table 7. The pretest data of both measures were submitted to a one-way analysis of variance (ANOVA) to test for significant differences between the groups at the start of the experiment. No significant differences were found. For both the GJT and the FWRT, a 2 (Conditions) × 3 (Times of measurement) repeated measures ANOVA was run, comparing the scores of the EI and II groups over time. For both tasks, significant overall progress over time was observed, as indicated by the significant main effects for Time (for the GJT: $F(1.65, 106.95) = 13.38$, $p < .01$; for the FWRT: $F(2, 130) = 12.84$, $p < .01$). An interaction between Time and Instruction was observed only for the GJT, $F(1.65, 106.95) = 3.39$, $p < .05$: EI produced higher scores. For the FWRT, there was no interaction between Time and Instruction.

Post hoc analyses showed that the main Time effect for the GJT was mainly due to the progress made by the EI group. Paired samples *t*-tests revealed that the

Table 7 Means and standard deviations for DoCs GJT and FWRT scores

Measure	Instruction	*n*	T0 *M (SD)*	T1 *M (SD)*	T2 *M (SD)*
GJT	EI	32	2.44 (2.08)	3.50 (1.93)	3.78 (1.98)
	II	35	1.94 (1.78)	2.06 (2.18)	2.49 (2.02)
Free written response	EI	32	5.44 (4.79)	7.53 (5.91)	7.25 (4.79)
	II	35	5.09 (4.72)	6.66 (4.54)	7.14 (4.37)

Note. DoCs = degrees of comparison; GJT = grammaticality judgment test; FWRT = free written response task.

Table 8 Post hoc analyses exploring the interaction between Time and Instruction for the grammaticality judgment test

Time of measurement	Observed difference[a]	F	Effect size (in d)
T0	EI = II	—	—
T1	EI > II	$F(1,65) = 8.14$	0.70
T2	EI > II	$F(1,65) = 7.01$	0.64

[a]Significant differences at alpha = .05 and using the Bonferroni adjustment are indicated by > or <.

only significant progress was made by the EI group between T0 and T1, $t(31) = -3.14$, $p < .01$. One-way ANOVAs were used to investigate the interaction between Time and Instruction. They revealed significant differences between the two groups at both T1 and T2, the EI group significantly outperforming the II group at both times. In Table 8, these analyses are reported along with the effect sizes when significant differences occurred. For the FWRT, there were no differences between the two conditions: Both progressed equally. Students had progressed primarily between T0 and T1. From T1 to T2, which actually spanned a longer time period, not much progress was made. This was tested by means of paired-samples t-tests. The factor Time was significant for T0 to T1, $t(66) = -3.98$, $p < .01$, but not for T1 to T2, which indicates that little progress was made after the instruction stopped. Overall, the size of the effect from T0 to T2 was moderate ($d = 0.41$) according to Cohen's (1988) standards.

The analyses presented here provide no evidence for the first hypothesis. One reason may be that not all participants in the EI condition were able to benefit from the instruction sufficiently, obscuring any effects present. To investigate this, the analyses were rerun, this time excluding participants in the EI condition who did not obtain an explicit knowledge increase of 1 standard deviation or more on the GJT. In doing so, 14 of 32 subjects that had received EI about the DoCs were lost. The II group was not reduced. The statistical power was still sufficient to detect medium effect sizes and there were no significant differences between the groups at T0. The results were the same: no interaction between Instruction and Time for the FWRT.

Results for Verb-final

The analyses presented in this section pertain to the second hypothesis: EI about Verb-final in subordinate clauses (V-final) leads to better performance on the

Table 9 Means and standard deviations for V-final GJT and FWRT scores

	Instruction	*n*	T0 *M (SD)*	T1 *M (SD)*	T2 *M (SD)*
GJT	EI	35	2.17 (2.41)	3.86 (2.64)	3.63 (2.69)
	II	32	1.44 (1.72)	2.44 (2.26)	3.19 (2.47)
Correct use	EI	35	5.94 (5.75)	7.03 (5.52)	7.54 (4.90)
	II	32	6.09 (5.42)	8.19 (5.39)	9.25 (5.94)

Note. V-final = Verb-final; GJT = grammaticality judgment test; FWRT = free written response task.

FWRT compared to II. Table 9 presents mean scores and standard deviations on the GJT and the FWRT for V-final. One-way ANOVAs were used to assess for both measures whether the groups differed significantly at T0. No significant differences were observed. For both tasks, a 2 × 3 repeated-measures ANOVA was run to compare the progress made by the two instruction groups. Main effects for Time were observed for both the GJT and FWRT (for the GJT: $F(2, 146) = 22.49$, $p < .01$; for the FWRT: $F(1.74, 113.35) = 13.31$, $p < .01$), but there were no interactions between Instruction and Time, meaning that the instruction led to significant and equal progress in both conditions for both tasks. Excluding T2 from the factor Time, thus increasing the subject sample to 81, led to the same result. The main effects found for both tasks were further explored by means of paired-samples *t*-tests. For the GJT, significant progress was made between T0 and T1, $t(66) = -5.67$, $p < .01$. Any progress made between T1 and T2 was not significant. This is an indication that little additional progress was obtained after the subjects stopped receiving instruction. The size of the effect from T0 to T2 was large ($d = 1.40$). For the FWRT, progress between both T0 and T1, $t(66) = -3.56$, $p < .01$, was significant. Progress obtained between T1 and T2, $t(66) = -1.92$, $p = .06$, just missed significance. The size of the effect from T0 to T2 was large ($d = 1.04$).

For subordinate clauses, the analyses were also rerun after exclusion of participants in the explicit condition that did not obtain an explicit knowledge increase of 1 standard deviation on the GJT or more. The EI group was reduced by 20; the II group was not reduced. At T0, there were no significant differences between the groups. The results for the 2 × 3 ANOVA were slightly different: There was a significant interaction between Instruction and Time for the GJT: the EI group being better on this task than the II group, which obviously was the result of our manipulation. For the FWRT, there was still no interaction between Time and Instruction.

Table 10 Distribution of participants across the conditions according to L1 similarity

	L1 expression of the degrees of comparison			L1 basic word order		
	Morphological	Periphrastic	Total	SVO	SOV	Total
Group 1 (EI in DoCs)	15	13	28	14	14	28
Group 2 (EI in V-final)	12	22	34	21	13	34
Total	27	35	62	35	27	62

Note. EI = explicit instruction; DoCs = degrees of comparison; V-final = Verb-final.

Effects of Instruction and L1 Similarity

The analyses in this section pertain to the third hypothesis: The success of EI and II depends on the interaction between the target structures and the way their meanings are expressed in the participants' L1s. For the DoCs, the participants were divided into groups according to whether their L1s encodes comparison morphologically or periphrastically. For V-final, the participants' L1s were coded for basic word order (SOV or SVO). Table 10 shows that a division according to L1 similarity led to a fairly even distribution across the groups. In order to achieve sufficient statistical power, 26 subjects per group are needed to detect large main effects and 13 subjects per group are needed to detect large interaction effects. One-way ANOVAs were used to check for differences between the groups on the GJT and FWRT scores at T0. For the DoCs, scores on both the GJT and the FWRT differed significantly. Participants with similar L1s already had more knowledge. For the V-final scores, there were no differences. Because of these results, gain scores rather than absolute scores were used for the present analyses. These gain scores reflect how much progress students obtained from moment to moment. They were effectively operationalized as the standardized residuals of pairs of observations. Gain scores were calculated for both structures and both measures from T0 to T1 and from T0 to T2.

With Instruction and L1 Similarity as two-level between-subjects factors, 2×2 univariate ANOVAs were used to test the hypothesis that the effect of instruction depends on the L2 learners' L1. This model did not explain a significant amount of variance for any of the computed gain scores, except for the DoCs on the FWRT. For both gain from T0 and T1 and from T0 to T2, the model was significant, although the amount of variance explained was largest for T0 to T1 (for T0 to T1: $R^2 = .21$, $F(3, 62) = 6.40$, $p < .01$; for T0 to T2:

Table 11 Means and standard deviations for the gain score on the free written response task from T0 to T1 for the degrees of comparison

	Morphological		Periphrastic	
	n	*M (SD)*	*n*	*M (SD)*
EI	15	0,92 (1,34)	13	−0,57 (0,57)
II	12	−0,09 (0,99)	22	−0,02 (0,72)

Note. EI = explicit instruction; II = implicit instruction.

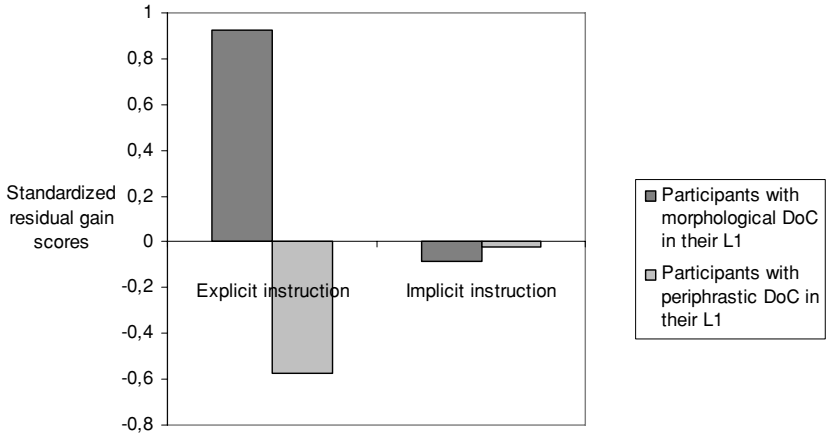

Figure 1 Standardized gain scores on the FWRT task from T0 to T1 for the DoCs, broken down by instruction condition and L1 similarity. *Note*. A gain score of zero reflects average rather than no gain.

$R^2 = .12$, $F(1, 62) = 3.48$, $p < .05$). There was no main effect for instruction; however, a main effect of L1 similarity was observed, $F(1, 61) = 8.61$, $p < .01$, as well as an interaction between Instruction and L1 similarity, $F(1, 61) = 10.25$, $p < .01$. Table 11 presents the descriptive statistics for the gain score on the FWRT from T0 to T1. Figure 1 illustrates the results graphically. To exclude the possibility that these results are due to differences between the groups at T0, we reran the analyses with the T0 scores as covariate. This made no difference to the results.

One-way ANOVAs were used to explore the observed between-subjects effects further. In the implicit condition, no significant differences were observed between participants with similar (descriptive) or different (periphrastic) ways of expressing degrees of comparison in their L1s in comparison to Dutch. However, in the explicit condition, participants with similar L1s made more

progress than participants with different L1s, $F(1,26) = 14.07$, $p < .01$. The size of the effect was large ($d = 1.46$). If students with similar L1s are compared across conditions, then students in the explicit condition made more progress than students in the implicit condition, $F(1,25) = 4.76$, $p < .05$. The size of this effect was large ($d = 0.86$). If participants had different L1s, students in the explicit condition made less progress than students in the implicit condition, $F(1,33) = 5.49$, $p < .05$. The size of the effect was also large ($d = 0.84$). This means that students with L1s that also express the DoCs by means of morphological suffixation did significantly better on the FWRT when they had received EI. EI seemed to have an inhibitory effect when the participants' L1s expressed the DoCs by means of periphrasis.

Discussion

This study was designed to test the hypothesis that EI can improve grammatical correctness in a FWRT more effectively than II. Care was taken to keep the exposure to the target structures in the explicit and implicit conditions equal: Even the sentences in which the structures appeared were literally the same but processed for different purposes. The results suggest that EI can be more effective than II, depending on the nature of the target structure and on the L2 learners' L1.

If the GJT results are considered, the findings of this study largely support the hypothesis that EI is more effective than II. For the DoCs, an interaction was found between Instruction and Time on the GJT: The EI group had performed significantly better. The results for the DoCs confirm previous findings. DeKeyser (1995) and Robinson (1996), for example, also found EI to be more effective than II for reliable and simple structures. In a similar vein, De Graaff (1997) found EI to work better for morphological structures. For the V-final structure, the observed interaction between Instruction and Time on the GJT was absent. Both conditions did significantly better at T1: The EI group simply did not have an advantage over the II group. Perhaps the EI about subordinate clauses was ineffective. The instruction may have been too brief, although many studies prove that brief instruction can lead to significant knowledge gains (e.g., Day & Shapson, 2001; Jourdenais, Ota, Stauffer, Boyson, & Doughty, 1995; Muranoi, 2000; Robinson, 1996; Salaberry, 1997). Another explanation could be that particular features of the target structure make it complex, too complex to be instructed explicitly, which would agree with De Graaff's hypotheses and findings that II was more effective for the syntactic structure.

At first sight, this study did not find any differences between EI and II in performance on the FWRT for both target structures. Both instruction conditions led to significant knowledge gains on the FWRT for both structures. Even when the results were manipulated to create large differences between the groups in explicit knowledge (as measured by the GJT), there were no overall differences in performance between the two conditions on the FWRT. This finding implies that implicit learning will occur as long as relevant information is processed, irrespective of how the information is processed and agrees with the notion of concomitant implicit learning during explicit instruction (Hulstijn, 2002). However, upon closer inspection, the picture was slightly more complex. For one of the two structures, the hypothesis that EI is more effective than II was supported if the participants' L1 was taken into account: EI was found to be more effective for those learners whose L1s express the DoCs similarly to the Dutch DoCs. Moreover, EI seemed to have an inhibitory effect for those learners whose languages had different ways to express DoCs.

There are several explanations for why L1 influence was observed for the DoCs only. Again, the quality of the instruction may have caused this result. If the EI about V-final was ineffective, it could not have had a facilitative effect. However, another explanation may be found in the differences that exist between the two structures. Most, if not all, languages have means to express the DoCs. In Dutch, they are realized morphologically and sometimes periphrastically. Other languages similarly either use morphological marking, periphrasis, or both. This may have enabled language learners to compare and contrast the structure with their L1. In addition, the structure carries a clear and straightforward meaning. Learning the structure requires making form-meaning connections and is supposed to be aided by a degree of explicitness (e.g., Doughty, 2001). In contrast, the V-final structure required in Dutch subordinate clauses is quite meaningless semantically. It should be noted that the structure is also quite unique. Although some Germanic languages, such as German and Frisian, also possess the structure, in most languages VO/OV differences in main and subclauses do not exist. Because subordination affects basic word order, it was decided to code the participants' L1s for basic word order. This may have been a contrast unrelated to the success of EI about the V-final structure. Another explanation might be that the instruction was improperly timed; perhaps the students were not sufficiently proficient or not developmentally ready. We will return to this issue briefly below.

The large difference in gain scores on the FWRT for the DoCs between students with similar or different ways of expressing degrees of comparison in their L1 provides evidence for an interaction between instruction and DeKeyser's

(2003) notion of subjective difficulty. Apparently, the DoCs were more difficult for a particular group of learners: those whose L1s express the DoCs periphrastically. The results support the hypothesis that there is an interaction among instruction, the structure's objective properties (does it involve an arbitrary form-meaning connection or must it be learned through association?), and its subjective properties. In this study, EI was found to affect free response task performance only (a) when the structure requires making a form-meaning connection and (b) when the learner's L1 expresses the intended meaning in a similar fashion.

The most surprising finding perhaps was that students with different ways of expressing DoCs in their L1s were found to be at a disadvantage in the EI condition in comparison to students with different L1s in the II condition, again for the DoCs only. In other words, students with different L1s seemed to be affected negatively by the EI. The SLA literature does not readily explain this finding. Part of the definition of explicit knowledge given by R. Ellis (2004) was that it can be incorrect or incomplete. The distance between the DoCs in the L1 and L2 may have been too large for participants to fully grasp the EI, leading to imperfect explicit knowledge and virtually no gains in free response task performance, whereas students in the implicit condition with different L1s did progress. Another explanation for this effect may lie in the nature of the EI. The EI in this study involved negative evidence regularly when the learners had to choose the correct form from three alternatives. In the implicit condition, the same input sentences were used in comprehension exercises, but always in their correct form. This may have created unequal opportunities to implicitly acquire the DoCs. Although possible, this explanation is less plausible, because a similar effect should have occurred for learners in the explicit V-final condition.

The fact that the amount of exposure to the target structures was equal in both conditions allows for the explanation that the effect of EI on the use of the DoCs in the FWRT was caused by explicit knowledge differences between the two conditions. These results, then, could be evidence of L2 learners "leaning on declarative crutches" (DeKeyser, 1998, p. 49). However, these data cannot be taken as evidence for an interface between explicit and implicit knowledge. They are congruent with all positions in the debate. Krashen (1985, 1994), who adopted a noninterface position, allowed for the possibility that L2 learners monitor their language use in certain nondemanding circumstances. The results may also be an indication of L2 learners actively and consciously applying explicit knowledge in a free response task (e.g., DeKeyser, 1998). Or the L2 learners in this study may have acquired the structure implicitly more effectively because of their explicit knowledge (e.g., Skehan, 1998). The exact

nature of the relationship between explicit and implicit knowledge remains unclear, not in the least because the GJT cannot be considered a pure test of explicit knowledge. It does not prohibit language learners from making use of their implicit knowledge. In fact, the finding that both conditions demonstrated equal progress on the GJT for the V-final structure suggests that participants reverted to their procedural knowledge in the absence of explicit knowledge. However, this study also provides some evidence for R. Ellis's (2004, 2005) claim that explicit and implicit knowledge can be separately tested. The fact that there was a difference in impact—for the DoCs at least—of EI on GJT and FWRT performance suggests that the tasks partly measure different constructs and that explicit and implicit knowledge can, to some extent, be measured independently. In addition, the correlations between the GJT and the FWRT were consistently moderate.

The interpretation of FFI research is complicated by the many factors that may have contributed to the results. The outcomes have been considered in the light of the research design used, the tests used to measure progress, and the objective and subjective qualities of the target structures. However, other factors may have affected the results as well. For example, the participants in this study were rather young, between the age of 12 and 18. A tremendous amount of literature has been devoted to age and explicitness of SLA (e.g., DeKeyser, 2000; Doughty, 2003; Muñoz, 2006; Robinson, 1997). The learners in this study may have been too young to optimally benefit from EI or young enough to learn implicitly from the input. Another aspect that may have had an impact is the fact that the acquisition of Dutch features of grammar was studied in a setting of immersion. R. Ellis (2002) has claimed that the availability of target structures in the everyday environment may be crucial to the success of instruction. Other factors that may have been of influence are the setting, the intensity, the modality, and the timing of the instruction. EI in this study constituted drill-and-practice exercises, but whether these results obtain in more communicative, so-called focus-on-form (FonF) settings (R. Ellis, 2001; Norris & Ortega, 2000) remains to be seen. Another feature of the instruction was that it was rather short; it consisted of no more than a few hours. Perhaps EI needs to be more intensive to be effective. Additionally, timing may also have influenced the results. It has been suggested that at least some parts of language may be subject to fixed orders of acquisition (R. Ellis, 1997; Lightbown, 1998; Pienemann, 1988) and that instruction will only work if it is provided when learners are developmentally ready to receive it. Finally, it remains to be seen whether the results can be generalized to oral free response tasks. As pointed out earlier, studies by VanPatten and Sanz (1995) and Day and Shapson (2001)

indicated that this is not necessarily so. The influence these factors may have had can only be disentangled by conducting similar studies in different settings.

There are several pedagogical implications to this study. If one can take the results of one study on one structure as evidence, this study suggests that EI may be disruptive for L2 learners whose L1 realizes the target structure differently. In addition, this study supports the notion that L1 influence is structure-specific rather than language-specific (Odlin, 2003). Teachers would need to figure out how each target structure relates to the L1 of their students. This will be difficult if all language learners in a course have the same language background and will be impossible if they all have different L1s. If there is one conclusion that can be drawn from FFI research, it is that instruction works (Lightbown, 2001; Long, 1983; Norris & Ortega, 2000). However, FFI research has not provided much evidence yet that EI should be preferred over II. Although replication and extension of these results is needed, this study suggests that EI may not be helpful in certain circumstances. Therefore, until more is known about when EI works, the safest option seems to be to implicitly intensify the input by offering structures in realistic and functional settings.

Revised version accepted 1 January 2009

Notes

1 A study by Ellis, Loewen, and Erlam (2006) also compared explicit and implicit types of instruction in relation to measures of explicit and implicit knowledge. The study was not included in this review because they used an oral imitation task as a measure of implicit knowledge based on R. Ellis (2004, 2005) rather than a free response task. Performance on the oral imitation task was positively affected by EI.

2 Given our design of random assignment to the treatment groups, we could also have presented analyses of covariance using pretest scores as covariate. We found the results were the same. We opted to present repeated-measures analyses for some of the analyses because they allow us to include both posttests in one analysis, in this way providing information about when most gain was obtained.

References

Andringa, S. J. (2005). *Form-focused instruction and the development of second language proficiency*. Unpublished doctoral dissertation, University of Groningen.

Authorware. (1999). San Francisco: Macromedia.

Bialystok, E. (1989). Psycholinguistic dimensions of second language proficiency. In W. Rutherford & M. A. Sharwood Smith (Eds.), *Grammar and second language teaching: A book of readings* (pp. 31–50). New York: Newbury House.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition, 16*(2), 157–168.

Carroll, J. B., & Sapon, S. M. (1959). *Modern languages aptitude test*. New York: Psychological Corporation.

Chaudron, C. (2006). Some reflections on the development of (meta-analytic) synthesis in second language research. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 323–339). Amsterdam: Benjamins.

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

Day, E. M., & Shapson, S. M. (2001). Integrating formal and functional approaches to language teaching in French immersion: An experimental study. *Language Learning, 51*(Suppl. 1), 47–80.

De Graaff, R. (1997). *Differential effects of explicit instruction on second language acquisition*. Unpublished doctoral dissertation, Vrije Universiteit Amsterdam.

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

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing in L2 grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42–63). Cambridge: Cambridge University Press.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*(4), 499–533.

DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Oxford: Blackwell.

DeKeyser, R. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning, 55* (Suppl. 1), 1–25.

Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206–257). Cambridge: Cambridge University Press.

Doughty, C. (2003). Instructed SLA: Constraints, compensation and enhancement. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 256–310). Oxford: Blackwell.

Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197–261). Cambridge: Cambridge University Press.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*, 305–352.

Ellis, N. C. (2006). Meta-analysis, human cognition, and language learning. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 301–322). Amsterdam: Benjamins.

Ellis, N. C., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics. *Applied Linguistics, 27*(4), 558–589.

Ellis, R. (1994). A theory of instructed second language acquisition. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 79–114). London: Academic Press.

Ellis, R. (1997). *SLA research and language teaching*. Oxford: Oxford University Press.

Ellis, R. (2001). Investigating form-focused instruction. *Language Learning, 51*(Suppl. 1), 1–46.

Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? A review of the research. *Studies in Second Language Acquisition, 24*, 223–236.

Ellis, R. (2004). The definition and measurement of explicit knowledge. *Language Learning, 54*(2), 227–275.

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

Ellis, R. (2007). Explicit form-focused instruction and second language acquisition. In B. Spolsky & F. M. Hult (Eds.), *The handbook of educational linguistics*. Oxford: Blackwell.

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

Goldschneider, J. M., & DeKeyser, R. (2001). Explaining the "Natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning, 51* (1), 1–50.

Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general proficiency. *Language Teaching Research, 2*(1), 1–23.

Harley, B. (1989). Functional grammar in French immersion: A classroom experiment. *Applied Linguistics, 10*(3), 331–359.

Housen, A., & Pierrard, M. (2005). Investigating instructed second language acquisition. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition*. Berlin: Mouton de Gruyter.

Hulstijn, J. H. (1999). *Vaardigheid zonder kennis?* Amsterdam: Vossiuspers AUP.

Hulstijn, J. H. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research, 18*(3), 193–223.

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Studies in Second Language Acquisition, 27*, 129–140.

Hulstijn, J. H., & de Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review, 11*, 97–112.

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

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. Lincolnwood, IL: Laredo Publishing.

Krashen, S. (1994). The input hypothesis and its rivals. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45–77). London: Academic Press.

Larsen-Freeman, D. (1995). On the teaching and learning of grammar: Challenging the myths. In F. R. Eckman, D. Highland, P. W. Lee, J. Mileham, & R. Rutkowski Weber (Eds.), *Second langauge acquisition theory and pedagogy* (pp. 131–150). Mahwah, NJ: Erlbaum.

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. New York: Longman.

Lightbown, P. M. (1998). The importance of timing in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 177–196). Cambridge: Cambridge University Press.

Lightbown, P. M. (2001). Classroom SLA research and second language teaching. *Applied Linguistics, 21*(4), 431–462.

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

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practise. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15–41). Cambridge: Cambridge University Press.

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal, 82*(3), 338–356.

Muñoz, C. (Ed.). (2006). *Age and the rate of foreign language learning*. Clevedon, UK: Multilingual Matters.

Muranoi, H. (2000). Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. *Language Learning, 50* (4), 617–673.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantative meta-analysis. *Language Learning, 50*(3), 417–528.

Odlin, T. (2003). Cross-linguistic influence. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 436–486). Oxford: Blackwell.

O'Malley, J. M., Chamot, A. U., & Walker, C. (1987). Some applications of cognitive theory to second language acquisition. *Studies in Second Language Acquisition, 9*, 287–306.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit learning: Implications for bilingualism and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393–419). London: Academic Press.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: Benjamins.

Pienemann, M. (1988). Determining the influence of instruction on L2 speech processing. *AILA Review, 5*, 40–72.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search and instructed conditions. *Studies in Second Language Acquisition, 18*, 233–247.

Robinson, P. (1997). Individual differences and fundamental similarity of implicit and explicit adult second language learning. *Language Learning, 47*(1), 45–99.

Robinson, P. (2002). Learning conditions, aptitude complexes and SLA: A framework for research and pedagogy. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 113–133). Amsterdam: Benjamins.

Salaberry, M. R. (1997). The role of input and output practice in second language acquisition. *Canadian Modern Language Review, 53*(2), 422–451.

Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning, 54*(1), 35–78.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–64). Honolulu: University of Hawai'i, Second Language Teaching & Curriculum Center.

Schuurs, U. (1999). *De ISK-toetsen*. Arnhem: CITO.

Sharwood Smith, M. A. (1988). Consciousness raising and the second language learner. In W. Rutherford (Ed.), *Grammar and second language teaching: A book of readings* (pp. 51–60). New York: Newbury House.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Spada, N., & Tomita, Y. (2008). The complexities of selecting complex (and simple) forms in instructed SLA research. In A. Housen & F. Kuiken (Eds.), *Proceedings of the complexity, accuracy and fluency (CAF) conference*. Belgium: University of Brussels.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*(2), 1–46.

Van Den Bercken, J. H. L., & Voeten, M. J. M. (2002). *Variantieanalyse: De GLM benadering*. Groningen: Stenfert Kroese.

VanPatten, B. (1996). *Input processing and grammar instruction*. New York: Ablex.

VanPatten, B. (2003). Input processing in SLA. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary*. Mahwah, NJ: Erlbaum.

VanPatten, B., & Sanz, C. (1995). From input to output: Processing instruction and communicative tasks. In F. R. Eckman, D. Highland, P. W. Lee, J. Mileham, & R. Rutkowski Weber (Eds.), *Second language acquisition theory and pedagogy* (pp. 169–185). Mahwah, NJ: Erlbaum.

Williams, J., & Evans, J. (1998). What kind of focus and on which forms? In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15–63). Cambridge: Cambridge University Press.