

Learner uptake and acquisition in three grammar-oriented production activities

Hayo Reinders *Innovation in Language Learning and Teaching*

This study investigates the effects of three types of production activities on uptake (operationalized as correct suppliance of the target structure during the treatment) and acquisition of negative adverbs in English. It also investigates the relationship between uptake and acquisition. The three production activities included a dictation, an individual reconstruction and a collaborative reconstruction activity. Each of these asked participants to produce the target structure but differed in (1) whether the activity was completed individually or collaboratively; (2) the amount of text participants had to produce; and (3) their degree of complexity and cognitive demand. It was found that all three activities resulted in uptake with the collaborative reconstruction, the dictation activities resulting in greater uptake than the individual reconstruction activity. There was also an effect for the activities on acquisition (of grammatical items only), but no differential effect for any of the three types of activities. It was concluded that a production activity can lead to increased uptake, but not to increased acquisition, and vice versa. The results may help language teachers look beyond immediate performance on an activity as a measure of success, and make better-informed decisions about when to use what type of activities.

Keywords: *SLA, output, complexity, cognitive demand, uptake, reconstruction, dictation*

I Introduction

As has been well documented, the role of grammar in the language classroom has changed dramatically over the years, with a shift from grammar-based teaching to the use of more communicative approaches and, more recently, the use of tasks to encourage a focus on meaning, the use of real-world processes, and a clearly defined communicative outcome (R. Ellis, 2000). Focused tasks that are communicative but also afford opportunities to focus on form have also become popular (R. Ellis, 2003).

It is not always clear if, and how, such activities aid learning. Neither is it clear on what basis teachers choose an activity to use in class. This study aims to investigate the effects on the development of one grammatical feature in English (namely negative adverbs) of three different types of grammar-oriented

Address for correspondence: email@hayo.nl

© The Author(s), 2009. Reprints and permissions: <http://www.sagepub.co.uk/journalspermissions.nav>
10.1177/1362168809103449

production activities (dictation, individual reconstruction and collaborative reconstruction) that are commonly used in language classrooms. It does this by looking both at immediate performance (uptake) and longer-term acquisition. The results may help teachers to make better-informed decisions about what activities to choose when, and may help researchers understand how activities differentially affect the learning process.

II Key constructs

1 Task complexity and cognitive demand

The production activities used in this study differ in three important ways. First, they differ in whether they are completed individually or collaboratively. Second, they differ in the amount of text participants are asked to reproduce. Finally, they differ in their complexity and, as a result, in their degree of cognitive demand. The concept of task complexity is particularly important as it is likely to affect task performance as well as acquisition. Robinson (2001a; 2001b) presents a framework to distinguish the relative complexity of tasks. He looks at task features (complexity), learners' perceptions (difficulty) and the conditions under which tasks are completed to facilitate decisions on task sequencing. Task complexity (the aspect that most directly concerns the study presented here) in this model is the 'result of the attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner' (Robinson, 2001b, p. 29) and is dependent on the extent to which tasks direct resources or deplete resources. Task complexity is determined by task characteristics, such as:

- here-and-now: with tasks not referring to the immediate context being more difficult;
- single task demand: the more tasks, the more demanding;
- planning time: the less time the more demanding;
- prior knowledge: the less learners know about a topic, the more demanding the task; and
- many elements: tasks requiring selection of certain aspects of the information available are more challenging.

To investigate (amongst others) whether there was a differential effect for production on complex and simple tasks, Robinson (2001a) randomly assigned participants to one of two task conditions: the simple task required participants to give each other easy directions from a map of an area they were familiar with. In the complex task they had to give more extensive directions from a map of an unfamiliar area. Results showed significantly greater lexical variety and more accurate language use on the more complex task, showing that the complex task pushed learners' production. Robinson did not investigate acquisition.

Task complexity is only one aspect of tasks that affects acquisition. In the general learning domain levels of processing theory, Craik and Lockhart (1972) claim retention is dependent on the depth to which information is processed, with deeper processing resulting in greater retention. Levels of processing theory has been extensively criticized, notably for failing to specify how to determine and measure the required depth of processing and for not explaining – but merely describing – the process of learning. Even so, its general thrust that greater cognitive commitment (or more ‘elaborate processing’ as it was later called) is a characteristic of successful learning seems to hold. Research on second language acquisition has, for example, found beneficial effects for tasks that involve verbal rehearsal (N. Ellis and Sinclair, 1996), for task repetition (Bygate, 1996; Leow, 1998) and for active participation in classroom interaction as opposed to observation only (Mackey, 1999); it has also found differential effects for various types of memory processes, where the lower level processes such as those resulting in storage in verbatim memory were found to be less successful than those making connections with long-term memory (Williams, 1999). In sum, cognitive commitment appears to be closely related to ultimate attainment. In the words of Robinson (2001a): ‘the greater the *cognitive demands* of a task, the more they engage *cognitive resources* (attention and memory), and so are likely to focus attention on input and output, which will have *performance effects*’ (p. 305; emphasis in original). This study looks at the different cognitive demands of three types of production activities (see further below).

2 Uptake and acquisition

In this study I also looked at both uptake and acquisition. Acquisition was operationalized as improved performance on a timed and an untimed grammaticality judgement test from pre-test to post-test. Uptake was operationalized as correct suppliance of the target structure during completion of the treatments. The reason for investigating both uptake and acquisition was that several authors have argued for the importance of investigating not only the effects of a treatment on acquisition, or the final ‘product’, but also the ‘process’ of language learning. N. Ellis and Schmidt (1997) write:

SLA research aspires to understand acquisition, and acquisition results from dynamic processes occurring in real time. It is difficult to gain an understanding of learning and development from observations of the final state, when we have no record of the content of the learners’ years of exposure to language nor of the developmental course of their proficiencies. If we want to understand learning we must study it directly. (p. 146).

One area of study that has looked at these ‘dynamic processes’ is the research done on the effects of ‘uptake’, as an incorporation by the learner of (part of) the input into their own production, especially in the context of the effects of teacher feedback on language learners’ production. Teachers, naturally, want to

know to what extent their feedback has an effect on their learners' immediate classroom production as well as on their longer-term development. R. Ellis and Sheen (2006) point out that it is not clear if uptake contributes to acquisition. Neither is it clear that absence of uptake indicates an absence of acquisition.

If learners correct their original errors by incorporating the target forms from recasts, then they must have noticed these forms at some level. However, failure to repair the original errors cannot be taken as evidence of a failure to notice the target forms. Indeed, as Ohta's (2000) study of Japanese foreign language classrooms showed, learners' responses to recasts often take the form of private rather than social speech. Similarly, repair does not signal that a learner has identified the corrective function of a recast, as echoing the interlocutor's recast might simply constitute an acknowledgment of the propositional veracity of the interlocutor's statement; this is a reply designed to keep the channel of communication open (p. 590).

The studies mentioned above have given insight into the process of classroom learning, and especially the effects of negative feedback. The concept of uptake has, however, to the best of our knowledge not been used to investigate:

- the effects of different types of activities (i.e. without teacher intervention); or
- the effects of positive input as opposed to negative feedback.

Nonetheless, uptake is one indicator of which aspects of the input learners notice and process to the extent that they are able to reproduce them correctly.

According to R. Ellis and Sheen (2006), 'Ultimately, whether uptake facilitates acquisition must be determined empirically. It is unfortunate that, to date, there have been very few studies that have examined this issue' (p. 590). Looking at uptake as well as its relationship with acquisition allows us to determine the effects of varying types of activities, which may affect uptake and acquisition differentially. This information could allow teachers to make better-informed decisions to balance different demands, such as the motivational need to encourage successful task completion, and longer-term acquisition. This study therefore looks at both the immediate effects of the three grammar-oriented production activities described below and at their delayed effects or, put differently, it looks at their effects on both uptake and acquisition.

II Investigating production activities

1 The role of output

In this study three types activities were used:

- a dictation;
- an individual reconstruction; and
- a collaborative reconstruction.

All three of these involve reproduction of parts of the input and thus involve learner output. The role of output and its effects on learning has been investigated quite extensively. Swain has argued for the important role of learner production in learning. Her ‘output hypothesis’ (1985) developed from observations of Canadian immersion students who, despite years of receiving exposure to the second language, did not fully develop in particular certain grammatical aspects of the target language. Swain found that the immersion classes were characterized by a lack of opportunities for output and afforded few opportunities for ‘pushed output’, i.e. output that required them to ‘stretch’ their interlanguage. Many students were able to get by using communication strategies to get their meaning across and were not challenged to further develop their language. Swain suggested that by requiring learners to produce comprehensible output, they would be pushed to be more accurate and to pay attention to both form and meaning, and in so doing move from semantic to syntactic processing. In addition, Swain (1998) suggested that output would:

- induce noticing;
- allow for hypothesis formation and testing (see also R. Ellis and He, 1999; Pica, 1988); and
- give opportunities for meta-talk.

The effect of output on noticing was investigated in another article (Swain and Lapkin, 1995) in the context of a writing task with a think-aloud protocol. The authors found that learners do become aware of problems in their writing and engage in strategic thought processes to solve those problems.

Swain (1985) has suggested that output can also serve to help with developing automaticity (referred to as the fluency function of output). This seems to have been corroborated by research that has shown that producing the language results in the strengthening of connections in the brain, which directly aids the process of automatization (compare de Bot, 1992; 1996). Izumi, in several studies, investigated the effect of output on noticing (2002; 2003; Izumi *et al.*, 1999). He investigated if there was both an immediate and a delayed effect for output by using a written reconstruction task to measure noticing of English relative clauses, followed by a series of post-tests. He found that in comparison with a control group who received a receptive (meaning-focused) task only, output did have an effect on both noticing and learning. Similar results were found by Song and Suh (2008), who used two types of output tasks (a reconstruction or a picture-cued writing task) and found increased noticing and acquisition of the English past counterfactual conditional for participants who had completed these tasks, compared with those who had not. However, this was only true for production tasks; no difference was found between the production and the control group on receptive knowledge of the target structure.

2 *The role of working memory in production*

In activities involving reproduction of (part of) the input, obvious factors affecting retention are the amount of text learners are asked to produce and the amount of time between being exposed to the input and reproducing it (i.e. immediate vs. delayed recall). N. Ellis has conducted a number of investigations into different aspects of memory for second language learning and its effects on acquisition (compare N. Ellis, 1996; 2001; 2003; N. Ellis and Beaton, 1993; N. Ellis and Sinclair, 1996; N. Ellis and Schmidt, 1997). N. Ellis and Sinclair (1996) discuss the role of working memory in language acquisition and argue that a large part of language learning – not just vocabulary learning, but also syntax, discourse and other aspects of language – involves the learning of sequences. The authors hypothesized that working memory is necessary for sequence information to be retained short term, and that especially rehearsal encourages the commitment of information to long-term memory. To test this they conducted a study in which they aurally presented participants with recordings of sentences containing a particular Welsh grammatical structure (soft mutation). They instructed some participants to repeat all utterances, and prevented others from doing so by instructing them to count when listening to the input. (This pertained only to the exposure phase; participants were able to speak as they wished at other times.) The authors assessed learning with the help of a grammaticality judgement test, a test of metalinguistic awareness, and a speech production test. They found that participants prevented from rehearsal had impaired performance at:

- a) comprehension and translation, b) explicit metalinguistic knowledge of the detailed content of grammatical regularities, c) acquisition of the foreign language forms of words and phrases, d) accuracy in pronunciation, and e) some aspects of productive, but not receptive, grammatical fluency and accuracy. (N. Ellis and Sinclair, 1996, p. 234)

3 *The role of collaboration*

Performance on a production activity is likely to be affected by whether the activity is completed alone or with another learner. Working with someone else has been said to have a range of benefits and there is, for example, a substantial body of research on the benefits of output and interaction on language learning (see Long, 1996). Classroom activities commonly require learners to exchange information, work collaboratively and interact in the target language; these characteristics have been investigated quite extensively. Looking at collaborative learning, for example, Swain and Lapkin (2001) investigated the acquisition of French pronominal verbs through two different types of activities. Participants received a short lesson on the topic and were then assigned to either a dictogloss or a jigsaw activity. The dictogloss required participants to listen to a passage spoken by a native speaker twice, take notes, and then work together on its reconstruction. This task was expected to encourage

more attention to form as a result of the native speaker input. The jigsaw task required participants to work together in reconstructing a story based on pictures of which each of the two participants in a pair held half. This type of activity was expected to offer many opportunities for negotiation of meaning, hypothesized to be beneficial for learning. Audio-recordings were made of the interaction between the participants, and the reconstructions and stories were investigated for the number of language-related episodes and qualitative differences. In addition, a pre-test and post-test investigated learning of the pronominal verbs. The findings showed that the dictogloss task resulted in participants producing fewer language-related episodes, probably because it was less open-ended than the jigsaw task. However, it also resulted in greater accuracy and more complex language. Contrary to expectations, there was no significant difference in the degree to which the tasks led participants to focus on the formal aspects of the language, nor in the post-test scores. The authors suggest that the preceding lesson – and the fact that in both tasks the end result had to be written down – focused participants' attention on language form. This appears to be corroborated by the fact that the total number of language-related episodes was substantial.

III Comparing the three production activities

I now briefly consider the key characteristics of each of the three production activities investigated in this study.

1 Dictation

One long-standing and still common type of production activity is dictation. Considering its ubiquity it is surprising how few empirical studies have investigated dictation, and there are no studies we are aware of that have investigated dictation from a cognitive perspective or have looked at the relationship between memory processes and task completion. One of the few recent studies on the subject by Kiany and Shiramiry (2002) investigated the effects of dictation on the listening comprehension of Iranian elementary level learners of English as a foreign language. The control group engaged in listening practice and exercises from their course book for the duration of the term. The experimental group did some of the exercises from the book and in addition did a dictation task on 11 occasions. Using a pre-test–post-test design, they found that the experimental group had significantly improved their listening comprehension in comparison with the control group. The authors took this as evidence that, despite the focus of the dictation task on form, this could positively affect participants' understanding as well.

Dictation usually requires learners to remember short sentences or chunks of longer sentences only, whereas reconstruction activities typically require

learners to piece together an entire text they have previously read or heard. It may be expected that having to remember only a small amount of text, as in the dictation activity, will allow learners to more accurately reproduce the text; in other words, that greater uptake is likely to take place. On the other hand, in the shorter amount of time it takes to complete the dictation, there may be less opportunity to rehearse the input in working memory and – following N. Ellis and Sinclair’s (1996) findings – to commit it to long-term memory. One of the purposes of this study is to establish whether the dictation affected immediate performance, and whether this resulted in greater acquisition compared with the individual and collaborative reconstructions.

2 *Individual reconstruction*

The individual reconstruction activity requires participants to remember a relatively long stretch of text and to reproduce it from memory (although usually participants are allowed to take notes). This is a type of resource depletion in Robinson’s terminology (2001a) as participants need to allocate limited cognitive resources to memorization. Unlike the collaborative reconstruction, the activity is completed individually, which means that cognitive resources cannot be pooled. The complexity of the task may affect immediate performance (uptake). At the same time, as the task is difficult, it may encourage participants to commit greater resources to it and this may have an effect on acquisition compared with the easier dictation.

3 *Collaborative reconstruction*

The dictation and individual reconstruction activity, on the one hand, and the collaborative reconstruction activity, on the other, differ in whether they are completed individually or in pairs. It may be expected that the collaborative aspect of the collaborative reconstruction will result in greater uptake compared with the individual reconstruction activity. The study will establish whether this is the case and, if so, whether this also affects acquisition. Table 1 summarizes the main characteristics of the activities used in this study.

Table 1 Summary of characteristics of the production activities used in this study

	Dictation	Individual reconstruction	Collaborative reconstruction
Cognitive demand	–	+	+/-
Reproduction	chunk	whole	whole
Individual or collaborative	individual	individual	collaborative

IV The study

1 Research questions

This study investigates the following research questions:

- What are the effects of (1) dictation, (2) individual reconstruction, and (3) collaborative reconstruction activities on uptake of English negative adverbs?
- What are their effects on acquisition?

2 Design

Participants completed a timed and an untimed grammaticality judgement test (GJT) as a pre-test and were then randomly assigned to one of three treatment conditions: a dictation, an individual reconstruction, or a collaborative reconstruction activity. There were three separate treatment administrations, one week apart. Immediately following the final treatment, participants completed a post-test (a timed and an untimed GJT) and a delayed post-test one week after that (a timed and an untimed GJT). For logistical reasons there was no separate control group in the study (see below). Instead, performance on target and non-target items was compared. Table 2 shows the design of the study.

3 Participants

The participants were 28 students enrolled in a full-time general English course in a language school based in a large city in New Zealand. They volunteered to join the study for a compensation of approximately US\$8 per hour. Sixteen of the participants were female and 12 male. Sixteen were from East Asia (Japan, Korea, China), four from Switzerland and the remaining eight from an additional six countries. Participants had nine different first languages. Most of the participants had lived in an English-speaking country for less than six months.

The participants had been given an in-house placement test earlier in the year to determine their class level. The school considered upper-intermediate level students to be at the equivalent of level B2 of the European Framework

Table 2 Design of the study

	Negative adverbs (<i>n</i> = 28)
Dictation	9
Individual reconstruction	8
Collaborative reconstruction	11

(Council of Europe, 2006). Students were selected at the upper-intermediate level to make it more likely that they were developmentally ready to acquire negative adverbs, but without having yet done so.

4 *Target structure*

The target structure was negative adverbs. Negative adverbs are adverbs or adverbial structures that lead to inversion of subject and auxiliary, as in:

Seldom *had he* seen such a beautiful woman.

Rarely *had he* seen such a beautiful woman.

Hardly *had he* seen such a beautiful woman.

Corpus searches show that negative adverbs are relatively infrequently used. They are also acquired late (compare Long, 1996; Pienemann, 1989).

5 *Treatment*

There were three different treatment types in the study. Each treatment provided participants with oral input on general interest topics that contained the target structure (negative adverbs). There were a total of 36 exemplars of the target structure across the three treatments. Participants had been randomly assigned to receive either procedural instructions on how to complete the task only ($n = 15$), or procedural instructions plus a noticing instruction ($n = 13$).

The noticing instruction read:

Listen carefully and pay attention to where the auxiliary verb comes in each sentence.

For example in the sentence 'Rarely has so much rain fallen in such a short time' the auxiliary is 'has' and it comes before the subject of the sentence 'so much rain'.

A previous study (Reinders and R. Ellis, 2009) made use of the same dataset but looked specifically at the effects of the absence or presence of the noticing instructions. It found no differential effect on either uptake or acquisition for the noticing instructions; in other words, and contrary to expectations, the noticing instructions did not affect participants' ability to produce the target structure (uptake) or acquire it, compared with the procedural instructions only. As the instructions had no differential effect, for the purposes of this study, the results for both groups are presented together.

a Dictation: In the dictation treatment participants listened to short texts of about 60–70 words on a computer, during which they were not allowed to take notes. On the second listening they heard the passage part by part and typed in each part, as they heard it. Each part contained no fewer than eight and no more than 10 words. This treatment thus involved immediate

recall. For participants completing the dictation treatment the instructions were:

In this task you will listen to some short passages. First, you will hear each passage entirely. Next you will hear the passage again, but this time bit by bit. Type out each bit *exactly as you hear it*. You will now have a chance to practise this.

b Individual reconstruction: In the individual reconstruction, participants listened to a passage of about 60–70 words twice. They were allowed to take notes. Next, they were asked to attempt to reconstruct the passage by typing it in on the computer. While they were doing this, participants were asked to talk-aloud (Ericsson and Simon, 1993) and their talk-aloud was audio-recorded. This treatment thus involved delayed recall. Instructions for the treatment were in the form of a video demonstrating talk-aloud where a student could be seen conducting the task while a voice-over explained the procedure. The instructions were:

In this task you will listen to a passage twice. Next, type in the passage exactly as you heard it. You are allowed to take notes.

When you try to reconstruct the passage, I want you to talk aloud. By talk aloud I mean that you say out loud everything that goes through your mind. Just act as if you are alone in the room talking to yourself. If you stop talking for any length of time I will remind you to keep talking.

c Collaborative reconstruction: The collaborative reconstruction treatment was similar to the individual reconstruction treatment except that two participants were paired and were asked to reconstruct the text together. Their discussion was audio-recorded. This treatment thus also involved delayed recall. For participants completing the collaborative reconstruction treatment the instructions also included a video demonstrating how to do the task. The instructions were:

In this task you will listen to a passage twice. Next, type in the passage exactly as you heard it. You are allowed to take notes.

Talk with your partner and help each other type out the passage. If you stop talking for any length of time I will remind you to keep talking.

All participants completed three practice passages before the start of each treatment. The treatment itself involved completing four passages containing three target sentences (i.e. sentences containing negative adverbs) each, in each of the three treatment sessions. There were therefore a total of 12 exemplars for each of the three treatments, and thus a total of 36 exemplars across the three treatments.

6 Tests

A timed GJT was administered on three occasions: as a pre-test, an immediate post-test and a delayed post-test. This test consisted of 50 sentences,

20 of which contained the target structure (negative adverbs). Of these, 10 were grammatical and 10 ungrammatical sentences. The other 30 items consisted of sentences with distractor items (relating to adverb placement or the difference in form between adverbs and adjectives). At each test administration the order of the items was changed. In the test, sentences were shown on the computer screen and participants had to press the 'enter' key if they thought the sentence on the screen was correct, and the left-hand 'shift' key if they thought it was not. The keys were labelled with stickers indicating 'correct' and 'incorrect'. There were eight practice sentences during which the researcher was present to give clarification where needed. The tests were first trialled on native speakers and learners of a similar level as those in the study, in order to establish a time limit for each sentence. The time limit for each sentence was longer than the mean time taken by the native speakers on that sentence but shorter than that of the non-native speakers. The learners were given relatively more time on the earlier than the later items in the test. They were told that they might not be able to respond to all the items in time but that they should try to answer as many as they could. Although GJTs are not without their drawbacks (see Birdsong, 1989), because of the infrequency of negative adverbs it would have been difficult to create more authentic test types that would still allow for sufficient items to be tested. The timed GJT also seemed appropriate as it has been used as a measure of implicit knowledge (compare Han and R. Ellis, 1998), which seemed the most likely outcome of the (implicit and incidental) activities used in the study.

The reliabilities of the negative adverb items and the control items were assessed by means of Cronbach alpha and are shown in Table 3. The alphas ranged from a high of .928 to a low of .605.

The treatments and tests provided:

- a measure of uptake of the target structure; and
- a measure of acquisition.

7 Uptake

In this study, uptake was operationalized as the correct suppliance of the target structure during the three treatments. Results were coded by the researcher with suppliance of the target form judged to be correct as long as

Table 3 Reliability figures for the grammaticality judgement tests (Cronbach alpha)

	Pre-test	Post-test	Delayed post-test
Non-target	0.841	0.928	0.840
Negative adverbs	0.605	0.779	0.732

the participants inverted subject and auxiliary. Spelling and other errors not relating to the target structures were discounted. For example, to the sentence 'No sooner does there seem to be a solution then another problem arises', one learner responded:

no sooner is it solution ... and the other problem is the ice

Here the wrong auxiliary was chosen but the word order was correct so the sentence was scored as correct. However, any sentence starting with an adverb and followed by a subject were scored as 'incorrect'. Sentences with no auxiliary (e.g. *No sooner that I arrived...*) or without a subject (e.g. *No sooner had arrive*) were also scored as incorrect.

8 Acquisition

Acquisition scores were arrived at by totalling the number of correct judgements that the learners made on the GJT. Overall gain scores and also separate scores for the 10 grammatical and the 10 ungrammatical sentences were calculated as previous research has indicated that these measure separate constructs (R. Ellis, 2005; Hedgcock, 1993). To measure acquisition, gain scores from pre-test to immediate post-test, from pre-test to delayed post-test and from immediate to delayed post-test were calculated. The likelihood that participants encountered the target structure outside of the treatment sessions was deemed to be low because of the infrequency of the target structure. A questionnaire administered to the participants' classroom teachers confirmed this.

9 Control

There was no separate control group. The main reason for this was the practical constraint of testing in a language school. Ideally there would have been a control group but it proved to be difficult to recruit and retain participants in the study. Recruitment and testing were done in the school, which placed a burden on administrators and teachers and also reduced classroom availability. Learners' responses to the 30 items in the GJTs that did not contain the target structure were used as a point of comparison to examine to what extent the effects of the treatments were greater than pure practice effects.

10 Analysis

As participants in the study completed multiple treatments and tests, repeated measures analysis of variance (ANOVA) models were used to investigate group differences. For *post hoc* analyses the least significant differences (LSD) method was used. This method is considered liberal in that it compares

means for all possible data sources separately, rather than combined. Considering the fairly small number of data sources, and considering that the present study was exploratory, the use of LSD was deemed acceptable. For all statistical analyses the alpha level was set at .05. For effect sizes, Cohen's *d* values were calculated.

V Results

First the results for uptake are presented followed by those for acquisition.

1 Uptake

Table 4 shows the results for the three treatment types for uptake of negative adverbs. The numbers represent the percentage of correct suppliance of the target structure. These show a clear improvement for scores for all three treatment types, from time one to time two and from time two to time three. This time difference was found to be significant ($F(1,80) = .41.57, p < .001$). In other words, the treatments had an effect on uptake of negative adverbs. The results show that scores on the dictation treatment are the highest at treatment time three, and scores on the individual reconstruction treatment are clearly the lowest. Scores on the dictation treatment show the greatest improvement over time, followed by scores on the collaborative reconstruction treatment. Scores on the individual reconstruction treatment also improved, but not as much. An ANOVA (3 = treatment types) shows a significant effect for treatment type ($F(2,80) = 11.01, p < .001$). *Post hoc* analyses show:

- a significant advantage for the collaborative reconstruction treatment over the individual reconstruction treatment ($p < .001$); and
- a significant advantage for the dictation treatment over the individual reconstruction treatment ($p < .001$). The effect size for treatment type is small ($d = .115$).

Table 4 Uptake scores, by treatment type (percentage of correct suppliance of the target structure)

	Time 1		Time 2		Time 3	
	Mean	SD	Mean	SD	Mean	SD
Dictation ($n = 9$)	.154	.147	.462	.317	.638	.325
Individual ($n = 8$)	.028	.055	.145	.152	.333	.251
Collaborative ($n = 11^*$)	.219	.119	.477	.263	.590	.251

Notes: * There were 11 collaborative participants because one did not complete the post-tests and was eliminated from the dataset.

2 Acquisition

Next the descriptive results for acquisition are presented as raw scores (Table 5) and gain scores (Table 6). Table 6 shows the mean gain scores for negative adverbs:

- from pre-test to post-test;
- from pre-test to delayed post-test; and
- from post-test to delayed post-test, for the GJTs.

Gain scores from pre-test to delayed post-test on grammatical items are greater for both the collaborative reconstruction and the individual reconstruction treatments than for the dictation treatment. On ungrammatical items, the scores on the individual treatment are the only ones to improve from pre-test to delayed post-test, and from post-test to delayed post-test.

First, the differences between total gain scores on target and distractor items were compared by means of a 2 (negative adverbs/controls) \times 3 (gain scores) repeated measures ANOVA. This showed no statistically significant difference ($F(1,333) = 1.16, p = .28$). In other words, the treatments had no effect on acquisition of negative adverbs as measured by overall gain scores on the GJTs. However, for the grammatical items in the GJT the gain scores for the negative adverbs were significantly greater than for the control items

Table 5 Raw scores for negative adverbs and non-target items on the GJT (percentage of correctly judged sentences)

	Target				Non-target			
	Grammatical		Ungrammatical		Grammatical		Ungrammatical	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
<i>Pre-test:</i>								
Dictation ($n = 9$)	.400	.200	.333	.193	.644	.187	.194	.126
Individual ($n = 8$)	.412	.195	.287	.155	.675	.138	.150	.119
Collaborative ($n = 11$)	.363	.156	.309	.186	.636	.174	.181	.078
<i>Post-test:</i>								
Dictation ($n = 9$)	.577	.281	.411	.208	.600	.217	.400	.234
Individual ($n = 8$)	.600	.250	.287	.172	.800	.160	.225	.119
Collaborative ($n = 11$)	.572	.313	.409	.287	.636	.254	.263	.145
<i>Delayed post-test:</i>								
Dictation ($n = 9$)	.522	.315	.322	.268	.522	.238	.372	.204
Individual ($n = 8$)	.662	.333	.437	.297	.700	.311	.381	.231
Collaborative ($n = 11$)	.645	.254	.300	.134	.718	.199	.268	.16

Table 6 Gain scores for negative adverbs and distractor items on the GJT

	Target items				Control items			
	Grammatical		Ungrammatical		Grammatical		Ungrammatical	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
<i>Pre-test to post-test:</i>								
Dictation (<i>n</i> = 9)	.177	.356	.077	.148	-.044	.343	.205	.144
Individual (<i>n</i> = 8)	.187	.322	.001	.106	.125	.198	.075	.119
Collaborative (<i>n</i> = 11)	.209	.320	.100	.282	.001	.167	.081	.118
<i>Pre-test to delayed post-test:</i>								
Dictation (<i>n</i> = 9)	.122	.281	-.011	.271	-.122	.315	.177	.182
Individual (<i>n</i> = 8)	.25	.277	.150	.282	.025	.237	.231	.231
Collaborative (<i>n</i> = 11)	.281	.256	-.009	.304	.081	.256	.086	.165
<i>Post-test to delayed post-test:</i>								
Dictation (<i>n</i> = 9)	-.055	.316	-.088	.226	-.077	.198	-.027	.171
Individual (<i>n</i> = 8)	.062	.329	.150	.338	-.100	.370	.156	.277
Collaborative (<i>n</i> = 11)	.072	.438	-.109	.341	.081	.282	.004	.118

($F(1,165) = 9.71, p = .002$) with a medium effect size ($d = .48$). There was also a significant difference on the ungrammatical items ($F(1,165) = 4.49, p = .035$), but this was to the advantage of the control items.

Next, an ANOVA was performed to establish if there was an effect for treatment type on the grammatical items. There was no effect for gains from pre-test to post-test ($F(2,53) = .38, p = .69$), from pre-test to delayed post-test ($F(2,53) = 2.08, p = .13$), nor from post-test to delayed post-test ($F(2,53) = 1.01, p = .37$).

3 Summary

The following is a summary of the main results:

- The treatments had an effect on uptake of negative adverbs.
- The dictation group outperformed the individual reconstruction activity on uptake of negative adverbs.
- The collaborative reconstruction group outperformed the individual reconstruction group on uptake of negative adverbs.
- The tasks had an effect on acquisition of grammatical items only.
- There was no difference between the three tasks on acquisition.

VI Discussion

The two research questions asked what effects the three treatment types had on uptake and on acquisition of negative adverbs. Uptake was measured in terms of the learners' suppliance of the target structure (negative adverbs with subject–verb inversion) in three different reproduction activities (dictation, individual reconstruction and collaborative reconstruction).

1 Uptake

The results showed a significant effect for the treatments on uptake of negative adverbs. *Post hoc* analyses revealed a significant advantage for the dictation treatment over the individual reconstruction treatment and for the collaborative reconstruction treatment over the individual reconstruction treatment. As discussed in the literature review, it is likely that the dictation was easier than the individual reconstruction treatment due to the shorter texts that had to be remembered for a shorter period of time. The collaborative reconstruction treatment had the advantage that it was completed in pairs so that participants could work together to piece the text together. The individual reconstruction treatment was clearly the most difficult for participants to complete.

2 Acquisition

There was no overall effect for the treatments on acquisition. What could explain this? The treatments may have been unsuccessful in encouraging participants to process the input for anything other than dealing with the immediate demands placed by the treatments. It is possible that activities such as those used in the present study that are inductive and do not include rule presentation at any stage – nor offer corrective feedback or negative evidence – may not affect participants' knowledge of the target structure. Although the target structure was presented a total of 36 times over three treatments this may not have been sufficient, especially for the relatively complex grammar point used in this study. Williams put it like this:

If learning distributional rules is critically dependent upon the subjects initially paying attention to relations between elements in the input, then it follows that even the simplest rules might not be learned if the subjects for some reason fail to attend to those relationships. (Williams, 1999, p. 32)

The activities used in this study appear to have caused the participants to notice the target structure, but not the underlying rules governing the behaviour of those structures.

3 *The relationship between uptake and acquisition*

The above results show that there is a differential effect for the three treatments on uptake on the one hand and acquisition on the other. Whereas the dictation resulted in greater uptake than the individual and collaborative reconstruction treatments, there was no evidence of greater acquisition resulting from the dictation treatment. These results also show that although uptake may be indicative of learning, it is not necessarily so, corroborating previous suggestions by R. Ellis and Sheen (2006).

One major difference between the dictation treatment, on the one hand, and the individual and collaborative reconstruction treatments, on the other, is the extent to which they engage cognitive processes. The dictation only required participants to keep the sentences in short-term memory for a limited length of time; it was thus relatively simple (in terms of Robinson's framework; 2001a), and this affected immediate performance. 'The effects of complexity differentials should be revealed by the fact that the cognitively simpler, less resource-demanding task will involve a lower error rate [...] than the more complex tasks' (p. 30).

The reconstruction treatments, on the other hand, required participants to remember relatively long stretches of text, and to remember them for a considerably longer length of time. Also, in order to reconstruct the content of the listening passages participants would have had to process the input semantically, whereas the dictation task did not require this level of processing. Semantic interpretation requires the activation of prior knowledge, and thus long-term memory processes which are linked to acquisition (see Robinson, 2003). The individual treatment, where participants did not get help from a partner, probably required the greatest amount of processing, and it may therefore have been the most cognitively complex of the three task types. In terms of Robinson's framework (2001a; 2001b) reported above, the reconstruction treatments involved greater resource depletion. This made it the least likely to result in participants correctly supplying the target forms. However, this does not necessarily mean that the target feature was not noticed and that some type of internal connection was not made; it is, however, unlikely to show from the participant's production, and there is thus a clear limitation evident here in that a measure of uptake cannot reveal this.

There may also have been effects for the verbalization required by the reconstruction tasks. Verbalization (especially during the individual reconstruction task) may have made it harder to accurately complete the task as it increases the number of demands on the learner and is thus resource-depleting (cf. Robinson, 2001a). On the other hand, verbalization may have also increased participants' cognitive commitment to the activity (by directing participants' attention) and thus ensured that what was noticed during completion was remembered better. As Robinson (2003) states: 'The greater the *cognitive demands* of a task, the more they engage *cognitive resources*

(attention and memory), and so are likely to focus attention on input and output, which will have *performance effects*.' (p. 305; emphasis in original). The findings also sit well with previous research by N. Ellis (1996; see also N. Ellis and Sinclair, 1996), who found that participants encouraged to rehearse foreign language utterances perform better on subsequent (productive) tests than both silent controls and participants prevented from rehearsal by articulatory suppression.

The relationship between the uptake and acquisition scores was strongest in the case of the collaborative reconstruction group. The collaborative aspect of the treatment was a practical benefit for immediate performance. This seems to corroborate claims by Swain (1995) and others for the benefits of interaction.

4 Pedagogical implications

One clear outcome is that the three activities did not have a strong effect on uptake or acquisition of this relatively complex structure. It is possible that for such structures either more explicit types of instruction or more extended exposure may be needed to have an effect.

Another outcome was the differential effect the treatments had on uptake and acquisition. Whereas the dictation resulted in greater uptake, it did not lead to greater acquisition. The individual reconstruction treatment – and to some extent the collaborative reconstruction treatment – showed the opposite results with relatively low uptake, but not less acquisition. Successful performance on a classroom activity is thus not necessarily a good predictor of ultimate acquisition. Of course, the benefits for acquisition are not the only consideration for teachers. Activities that lead to better immediate performance may motivate learners more, for example.

VII Conclusions

This study has shown that the types of activities used here do not greatly benefit either uptake or acquisition. It has also shown that different activities can have one, immediate effect and also a different longer-term effect. There are, however, a number of limitations to this study. First, the sample size was small. Second, there was no separate control group. Finally, the total exposure time provided by the instruction was relatively limited. This study included multiple instances of the target feature to investigate the relationship between uptake and learning aspects of the L2, but it is likely that in many teaching contexts the use of certain structures would not be predetermined in such a way. It is hoped that future studies will continue to compare the effects of specific task types and look at both uptake and acquisition to investigate both the process and the product of language learning.

VIII References

- Birdsong, D. (1989). *Metalinguistic performance and interlinguistic competence*. New York: Springer.
- Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis and M. Willis (Eds.), *Challenge and change in language teaching* (pp. 136–146). Oxford: Heinemann.
- Council of Europe (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Craik, F.I.M. and Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–683.
- de Bot, K. (1992). A bilingual processing model: Levelt's 'speaking' model adapted. *Applied Linguistics*, 13, 1–24.
- de Bot, K. (1996). Language attrition. In P. Nelde and W. Wölck (Eds.), *Handbuch Sprach Kontakt*. Berlin: Mouton de Gruyter.
- Ellis, N.C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18, 91–126.
- Ellis, N.C. (2001). Memory for language. In P. Robinson and J. Richards (Eds.), *Cognition and second language instruction* (pp. 33–68). Cambridge: Cambridge University Press.
- Ellis, N.C. (2003). Constructions, chunking, and connectionism: The emergence of Second language structure. In C. Doughty and M. Long (Eds.), *The handbook of second language acquisition* (pp. 63–103). Oxford: Blackwell.
- Ellis, N.C. and Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology*, 46A, 533–558.
- Ellis, N.C. and Schmidt, R. (1997). Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition*, 19, 145–171.
- Ellis, N.C. and Sinclair, B. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology*, 49A, 234–249.
- Ellis, R. (2000). Task-based research and language pedagogy. *Language Teaching Research*, 4, 193–220.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- 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. and He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, 21, 285–301.
- Ellis, R. and Sheen, Y. (2006). Re-examining the role of recasts in second language acquisition. *Studies in Second Language Acquisition*, 28, 575–600.

- Ericsson, K. and Simon, H. (1993). *Protocol analysis: verbal reports as data*. 2nd edition. Cambridge, MA: MIT Press.
- Han, Y. and Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research*, 2, 1–23.
- Hedgcock, J. (1993). Well-formed vs. ill-formed strings in L2 metalingual tasks: Specifying features of grammaticality judgements. *Second Language Research*, 9, 1–21.
- Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis. *Studies in Second Language Acquisition*, 24, 541–577.
- Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the noticing hypothesis. *Applied Linguistics*, 24, 168–196.
- Izumi, S., Bigelow, M., Fujiwara, M. and Fearnow, S. (1999). Testing the output hypothesis: Effects of output on noticing and second language acquisition. *Studies in Second Language Acquisition*, 21, 421–452.
- Leow, R.P. (1998). The effects of amount and type of exposure on adult learners' L2 development in SLA. *Modern Language Journal*, 82, 49–68.
- Long, M.H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie and T. Bhatia (Eds.), *Handbook of research on second language acquisition*. New York: Academic Press.
- Ohta, A.S. (2000). Rethinking recasts: A learner-centered examination of corrective feedback in the Japanese language classroom. In J.K. Hall and L.S. Verplaeste (Eds.), *The construction of second and foreign language learning through classroom instruction* (pp. 47–71). Mahwah, NJ: Erlbaum.
- Pica, T. (1988). Interlanguage adjustments as an outcome of NS–NNS negotiated interaction. *Language Learning*, 38, 45–73.
- Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, 10(1), 52–79.
- Reinders, H and Ellis, R. (2008, forthcoming). The effects of two types of positive enhanced input on intake and L2 acquisition. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp and H. Reinders (Eds.), *Implicit and explicit knowledge in a second language*. Clevedon: Multilingual Matters.
- Reinders, H. and Ellis, R. (2009). The effects of two types of positive enhanced input on intake and L2 acquisition. In R. Ellis, S. Loewen, R. Erlam, J. Philp, C. Elder and H. Reinders (Eds.), *Implicit and explicit knowledge in a second language*. Clevedon: Multilingual Matters.
- Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press.
- Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring Interactions in a componential framework. *Applied Linguistics*, 21, 27–57.
- Robinson, P. (2003). Attention and memory during SLA. In C. Doughty and M. Long (Eds.), *The handbook of second language acquisition* (pp. 631–678). Oxford: Blackwell.

- Song, M. and Suh, B. (2008). The effects of output task types on noticing and learning of the English past counterfactual conditional. *System*, 36, 295–312.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook and B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: studies in honor of Henry Widdowson* (pp. 125–144). Oxford: Oxford University Press.
- Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty and J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–81). Cambridge: Cambridge University Press.
- Swain, M. and Lapkin, S. (1995). Problems in output and the cognitive processing they generate: A step towards second language learning. *Applied Linguistics*, 16, 371–391.
- Swain, M. and Lapkin, S. (2001). Focus on form through collaborative dialogue: exploring task effects. In M. Bygate, P. Skehan and M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 99–118). Harlow: Longman.
- Williams, J. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition*, 21, 1–48.