Form-focused instruction and the acquisition of tense by Dutch-speaking learners of English: Experimental studies into the effects of input practice and output practice

Ureel, J.J.J.

Citation for published version (APA):
CHAPTER 5
STUDY 2

A statistical analysis, properly conducted, is a delicate dissection of uncertainties, a surgery of suppositions
(M. J. Moroney)

5.1 Introduction

The insights gained from the data analyses for Study 1 were used as a starting point for the design for Study 2. As discussed in Chapter 4, Section 4.5, four changes were implemented in the setup of Study 2. By way of summary, these four changes are listed below:

1. A reduction of the number of treatments from three to two, resulting in an input-practice group and an output-practice group (and the omission of the input-only group),
2. A focus on pretest and posttest outcome measures during the practice sessions, resulting in grammaticality judgement (GJ) and selected response (SR) practice for the input-practice group, and constrained constructed response (CCR) and translation (TR) practice for the output-practice group,
3. An evenly distributed increase of the pretest and posttest items from 24 to 32 over the four outcome measures as a whole, resulting in two additional items per outcome measure,
4. An exclusive focus on tense-related questions during the pretest and posttest sessions.

Since many of the design-related and methodological features for Study 2 were identical with those found in Study 1, I will refrain from consistently pointing out the details related to any overlap between both studies. In the event of any large overlaps, the reader will be advised to consult the information described in detail for Study 1 in Chapter 4. For example, the focus in the Methods section in this chapter (Section 5.2) will be on the changes that were implemented for Study 2. Overlaps between Studies 1 and 2 will occasionally be highlighted but unless differences are explicitly discussed, the reader should view the setup of Study 1 as identical with the setup of Study 2.

Once again, the general research question for Study 2 was whether explicit FFI had any effect on the acquisition of the complex past/present perfect distinction (when used to locate bygone situations) by Dutch-speaking ESL learn-
For Study 2, the three research questions for Study 1 were formulated once again and are provided below by way of recapitulation.

**Research questions**

1. Does explicit form-focused instruction (FFI) have an effect on (Dutch-speaking) ESL learners' performance with regard to complex temporal FMU mappings (i.e., the past/present perfect distinction) in L2 English?

2. If explicit FFI does have an effect, can any overall differential effects be ascertained with respect to the specific type of instructional treatment (e.g., input practice, output practice)?

3. If differences between treatments can be ascertained, are the differential effects the same across all of the outcome measures? Or are there differences between receptive and productive outcome measures?

The research hypotheses accompanying the three research questions were identical with the ones formulated for Study 1. Once again, we did not include a no-treatment (control) group to address research question 1. We simply assumed that an effect of treatment would be visible on posttest performances.

**5.2 Method**

**5.2.1 Design**

For Study 2, the four-way mixed-design template which was used for Study 1 was re-used as a design template. However, one structural change was made. Instead of assigning three levels to the between-subjects variable treatment—as was the case in Study 1 (input only, input practice, output practice)—we assigned two levels to the treatment variable in Study 2 (input practice, output practice). In other words, we did away with the input-only level. The within-subjects variables were once again time, with two levels (pretest, posttest), modality, with two levels (receptive, productive) and outcome measure, with four levels (grammaticality judgement (GJ), selected response (SR), constrained constructed response (CCR) and translation (TR)). Consequently, Study 2 was drawn up using a four-way 2 (Treatment) x 2 (Time) x 2 (Modality) x 4 (Outcome Measure) mixed-design template. The within-subjects variables were nested in exactly the same way as in Study 1, with modality nested hierarchically under time, with outcome measure nested non-hierarchically under modality (see Figure 5.1). Consequently, it was once again inappropriate to conduct an analysis including all independent variables since the four outcome measures were split across the two modalities, with GJ and SR nested under the receptive level of the modality variable and CCR and TR nested under the productive level of the modality variable. As a result, we decided once again to carry out two separate analyses (see Section 5.3).
CHAPTER 5

153

dependent variable

1
continuous

independent variables

4
categorical

written performance on a 32-item tense-related test consisting of four separate 8-item outcome measures of either a receptive or a productive nature: (1) grammaticality judgment, (2) selected response, (3) constrained constructed response and (4) translation

Between-groups variable:
1) Treatment (2 levels)
   1) Input practice
   2) Output practice

Within-groups variables:
1) Time (2 levels)
   1) Pretest
   2) Posttest
2) Modality (2 levels)
   1) Receptive
   2) Productive
3) Outcome Measure (4 levels)
   1) Grammaticality judgment
   2) Selected response
   3) Constrained constructed response
   4) Translation

Figure 5.1. Design box for Study 2 (2 x 2 x 2 x 4 mixed design)

5.2.2 Participants
The same population of ESL learners as the one under investigation in Study 1 was used for the selection of a sample of participants for Study 2. The only difference was that the participants were new and had not participated in Study 1. Because there were students who had not passed the English Grammar course the year before and, consequently, had already taken part in Study 1, we decided to allow these students to take part in Study 2 but their data were not used to avoid the influence of confounding variables (e.g., carry-over effects). Any data reported below automatically exclude participants who were retaking the English Grammar course and, consequently, had already participated in Study 1.
Attrition. Initially, 88 participants took part in the pretest but the number of participants dropped to 79 during the four experimental sessions. This drop corresponded with an attrition rate of approximately 10%. Once again, the drop in participant numbers was the result of extraneous factors (see Chapter 4, Section 4.2.2, Attrition). For the analysis and discussion of the experimental data, the 10% attrition rate was taken into consideration. We also tried to avoid any ceiling effect from influencing the data by screening the pretest data and excluding—as we did for Study 1—any participants who scored more than 85% of the pretest items correctly. In practice, 85% on the pretest corresponded with a score of 28 or higher on a possible maximum score of 32. The data from eight participants were not taken into consideration as a result of this screening procedure. After taking into consideration the attrition rate and the screening procedure, 71 participants remained. The data from these 71 participants were considered in the descriptive and inferential data analyses.

Assignment to treatment groups. The participants were assigned to treatment groups in the same way as in Study 1. This resulted in random assignment to one of the two treatment groups as follows: 36 (input practice) and 35 (output practice).

ESL grammar course. The experimental sessions were once again conducted as part of the participants’ regular English Grammar course in the first year of their English bachelor programme. The only difference was that Study 2 took place in the academic year 2008-2009, which was a year after Study 1. All the other course-related information was the same as the information provided for Study 1 (see Chapter 4, Section 4.2.2, ESL grammar course).

ESL language proficiency. In line with the criteria used for Study 1, the participants’ overall level of proficiency for English was categorized as intermediate to upper intermediate. The categorization criteria which were used for Study 2 were the same as those used for Study 1.

Academic exposure to ESL English. As far as the academic exposure to ESL English was concerned, no changes were reported with the situation described for Study 1.

Academic exposure to other L2s. The foreign language combinations were once again dependent on the participants’ own choices within the constraints of the bachelor programme structure in use. The context in which the foreign language combinations were made was described in Chapter 4 so the reader is advised to consult the details reported there for any background information regarding the participants’ foreign language combinations (Chapter 4, Section 4.2.2, Academic exposure to other L2s). The descriptives for the 71 participants’ foreign language combinations found in Study 2 are presented in Table 5.1.
Table 5.1. Participants’ language combinations (Study 2)

<table>
<thead>
<tr>
<th>English</th>
<th>other L2(s)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>Spanish</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Chinese (Mandarin)</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Italian</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Russian</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>French</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>6</td>
</tr>
<tr>
<td>Subtotal</td>
<td></td>
<td>65</td>
</tr>
</tbody>
</table>

| yes     | French and Italian | 1 |
|         | French and Spanish | 4 |
|         | German and Russian | 1 |
| Subtotal|             | 6 |
| Overall total |            | 71 |

Of the 71 participants, 65 had selected English as part of an L2 combination consisting of two L2s. As far as instruction was concerned, these 65 participants received seven hours of ESL instruction on a weekly basis. For their other L2s in their L2 combinations, they also received seven hours of instruction on a weekly basis. As was the case for all of the other first-year participants, these 65 participants were expected to take Dutch language courses since Dutch was their L1 language. The remaining six participants had chosen English as part of an L2 combination consisting of three L2s. These participants received either six or seven hours of English on a weekly basis, which they combined with two other L2s and Dutch (as their L1 language).

**Gender and age.** Of the 71 participants, 21 were male and 50 were female. The participants ranged in age from 18 to 29 years and the mean age was 20 years ($SD = 2$, mode = 19, median = 19).

**L1 backgrounds.** As was the case for the participants in Study 1, the participants in Study 2 were asked about their L1 backgrounds and about which language(s) that they themselves regarded as their L1(s). Of the 71 participants, 70 indicated that they had one L1. Of those 70 participants, 68 participants had Dutch as their L1s. The remaining two participants’ L1s were Albanian and Polish. One participant indicated that she had two L1s, which were Dutch and Russian.

**Remuneration.** The same remuneration information as that provided for Study 1 was applied for Study 2.
5.2.3 Apparatus and materials

5.2.3.1 Apparatus

The participants completed all of the four experimental sessions in the same computer language lab that was used for Study 1. In addition, the infrastructure in the language lab was the same as that found in Study 1. The only difference were the updated computers and computer programs which the participants used for the experimental sessions. Since these updates were of no significance for the actual execution of the individual experimental sessions, no additional update-related information will be provided.

For Study 2, the computer specialist who was contracted for Study 1 was once again asked to oversee the programming of the software. Apart from some design-related changes (see Section 5.2.1), no fundamental changes were made to the programming of the experimental sessions. Technical support was again available throughout the entire duration of the four experimental sessions and no major computer-related and/or network-related problems were reported.

5.2.3.2 Materials

For a detailed description of the materials used for Study 2, the reader is advised to read Chapter 4, Section 4.2.3.2, which provides in-depth information about the materials used. Since most of this information did not change for Study 2, the description provided in this section will be limited to the changes that were made to the materials. The changes will be discussed in the following subsections: pretest, treatment theory, treatment practice and posttest. This subdivision reflects the setup of the experimental sessions in Study 2.

Pretest. The corpus of texts which was used for Study 1 was also used for Study 2. For information related to the creation of the corpus, see Chapter 4, Section 4.2.3.2, Pretest. However, instead of selecting 24 items for the pretest, we decided to increase the number of pretest items to 32 items. This increase allowed us to gather more data and to do so across the four outcome measures, with two items being added to every outcome measure. A second change in the pretest was the number of questions per text. Whereas in Study 1 we had decided to present the participants with one tense-related question and three additional questions related to other aspects of the texts (e.g., cultural references, grammar items not related to tense, lexis), we decided to do away with the questions which were not tense-related. The main advantage of asking only one (tense-related) question per text was that the pretest became less time-consuming and could be taken in one experimental session, which was beneficial for the internal validity of Study 2. In Study 1, the participants were given two experimental sessions to complete the pretest because they were asked to answer four questions per test item.

The types of tense-related questions that were asked were identical with the ones that were asked in the pretest for Study 1. For additional information on
the nature of the tense-related questions and for examples of such questions, see Chapter 4, Section 4.2.3.2, Pretest.

**Treatment theory.** No changes were made to the treatment theory. All of the participants received the exact same theoretical information. For detailed information on the theory session, see Chapter 4, Section 4.2.3.2, Treatment theory.

**Treatment practice.** For Study 2, there were two experimental groups—input practice and output practice—which received different forms of tense-related treatment practice. Although the two experimental groups were given names which are identical with two of the three treatments in Study 1, the nature of the forms of treatment practice in Study 2 were structurally different from the forms of treatment practice in Study 1. The seven texts that were used for the treatment practice in Study 1 were re-used (with the addition of one text to bring the number of texts to eight, which allowed for an even distribution of texts across the four outcome measures) but the nature of the practice was changed. In other words, the treatment practice reflected more closely the input/output distinction found for the pretest and posttest outcome measures in Study 2. In essence, the practice was geared towards the four outcome measures used for Study 2. This meant that the input-practice group received practice for grammaticality judgement (GJ) task types and selected response (SR) task types whereas the output-practice group received practice for constrained constructed response (CCR) task types and translation (TR) task types. The materials that were used for the practice sessions consisted of eight texts, which were identical for both instructional treatments. For the practice geared towards the GJ task types, the verb form of interest (i.e., either a past or a present perfect) changed colour (from black to blue) and appeared in bold in the text 40 seconds after the text had appeared on-screen. Subsequently, the input-practice participants received a question about the verb form in bold. The input-practice participants were then requested to select one of the two options (see Figure 5.2).

Upon confirmation of their answers, the input-practice participants received a slide containing metalinguistic feedback on the linguistic context in which both the finite verb form and other relevant temporal information available in the text were highlighted. The feedback provided to the input-practice participants in Study 2 was identical with the feedback provided to the input-practice participants in Study 1. For the treatment practice geared towards the SR task types, the input-practice participants received a text with a selected response task type (see Chapter 4, Figure 4.3). This was followed by a picture selection task with feedback. The output-practice participants received the same texts as the input-practice participants but with CCR task types and TR task types (see Chapter 4, Figures 4.4 and 4.5). Additionally, the output-practice participants also received the picture selection task with feedback, which the input participants also received.
Louis Armstrong (1901-1971) was one of the greatest jazz trumpet players of the 20th century, but he ended his career though as a singer. Lips are not suited to vibrate against the mouthpiece of a trumpet day in and day out. Armstrong struggled with this fact all his life. Even at the height of his career in the 1930s, his lips had been ruined as a result of his energetic blowing technique. He was always treating his injuries and blisters using cream. Every time he put his instrument against his lips it felt as though a scorching poker was being held against them. At the end of a concert he was forced to swallow the constant stream of blood. There was a concert in 1935 during which the blood flowed down his shirt during his performance. In the end, Armstrong was forced to cancel so many performances that his manager sued him for breach of contract.

**Question**

In standard, formal present-day English, the verb form in bold in the text above is considered . . .

- grammatically correct
- grammatically incorrect

**Figure 5.2. Example text and question for input-practice group**

**Posttest.** One week after completing the treatment sessions, the participants were asked to take part in an unannounced posttest. The setup of the posttest in Study 2 was identical with the setup of the pretest. Once again, the difference was the actual nature of the posttest items. Whereas the pretest consisted of 32 new items, the posttest consisted of 16 ‘old’ items (i.e., items which had been recycled from the four pretest outcome measures, with four recycled items per outcome measure) and 16 ‘new’ items, which were new for all of the participants.

**5.2.4 Procedure**

Study 2 was organized in the month of March of the academic year 2008-2009. It consisted of four experimental sessions, which were divided as follows: pre-test (one session), treatment theory (one session), treatment practice (one session), posttest (one session).

Once again, the participants were invited to enrol before the beginning of the first experimental session. The participants were told that the experimental sessions would focus on an aspect of English grammar but no further information was provided to prevent the participants from preparing for the experimental sessions.
During the first experimental session, the participants were informed that they would be taking a test and that any test-related information would be provided to them on-screen. The instructions that the participants received for the pretest in Study 2 were almost identical with the instructions given to the participants in Study 1. The only difference was that the participants were instructed to complete the entire pretest in one experimental session. One week later, the participants received explicit information about the ESL target features under investigation. A week after receiving this information the participants took part in a practice session, which was different depending on which of the two treatments the participants had been assigned to. Since there were no changes for these two sessions between Studies 1 and 2, no further procedural information will be provided here. A week after completing the practice session, the participants were invited to take a posttest. Like the pretest, the posttest consisted of 32 items. The selection of the 32 posttest items for Study 2 was completed in the same way as the selection of the 24 posttest items for Study 1 (see Chapter 4, Section 4.2.4).

5.2.5 Coding and scoring
No changes were made to the coding and scoring procedure. In other words, the procedures explained for Study 1 in Chapter 4, Section 4.2.5, were replicated and used for Study 2.

5.2.6 Statistical analyses
With regard to the statistical analyses, the same logic that was applied to the design in Study 1 was applied to the design in Study 2. Instead of running a 2 (Treatment) x 2 (Time) x 2 (Modality) x 4 (Outcome Measure) analysis, we decided to run two separate analyses (see Section 5.3). For any other details concerning the statistical analyses, the reader is advised to read Chapter 4, Section 4.2.6.

5.2.7 Validity and reliability
Since we used the same corpus of texts for the selection of pretest and posttest items, the information provided in Chapter 4, Section 4.2.7, on the validity of the items also applies to the texts which were used for Study 2. Reliability was once again measured using Cronbach’s alpha coefficient (α). The Cronbach’s alpha coefficients for the pretest and posttest in Study 2 were .63 and .73 respectively.
5.3 Results

No qualitative data were collected for Study 2. Consequently, the results reported in this section are all based on the quantitative data, which were collected throughout the pretest and posttest sessions.

Once again, analyses for both main effects and interaction effects were carried out at all of the three levels of the experimental design as indicated in Figure 4.9 in Chapter 4.

As mentioned in Section 5.2.2, eight students scored 28 or higher on the pretest. In other words, they answered more than 85 per cent of the pretest items correctly. As was the case for Study 1, we took these students to be familiar with the ESL target structures under investigation and instruction was not likely to improve their performance significantly. Consequently, we conducted the two analyses on only those participants with pretest scores of 27 or lower, leaving 36 and 35 participants in the input-practice and output-practice groups respectively. We first conducted a three-way 2 (Treatment) x 2 (Time) x 2 (Modality) RM ANOVA, with treatment as the between-subjects variable and time and modality as the within-subjects variables. The descriptives are shown numerically in Table 5.2 and graphically in Figures 5.3, 5.4 and 5.5.
This analysis yielded a significant effect of time, $F(1, 69) = 131.060$, $p = .000$, $\eta^2_p = .655$. Performance on all 32 items showed a statistically significant, large increase from 20.32 in the pretest to 24.70 in the posttest, that is, an increase of slightly more than four points on a 32-point scale. No main effects of treatment or modality were obtained. In addition, most of the interactions were not significant. The only interaction which was significant was the Time x Modality interaction, $F(1, 69) = 9.319$, $p = .003$, $\eta^2_p = .119$ (see Figure 5.6). However, since the mean scores differ only marginally across time and modality, we attribute no importance to this finding in light of the research questions under investigation.
Figure 5.3. Performance on receptive pretest and posttest by treatment group (maximum scores of 16)

Figure 5.4. Performance on productive pretest and posttest by treatment group (maximum scores of 16)
Subsequently, we conducted a three-way $2 \times 2 \times 4$ (Treatment $\times$ Time $\times$ Outcome Measure) analysis. Numeric and graphic representations of the descriptives may be found in Table 5.3 and Figures 5.7, 5.8 and 5.9.
Apart from the main effect of time (the same effect as in the previous analysis, $F(1, 69) = 131.060, p = .000, \eta^2_p = .655$), a main effect of outcome measure was obtained, $F(3, 207) = 13.356, p = .000, \eta^2_p = .162$. There was one significant interaction, which was the Time x Outcome Measure interaction, $F(3, 207) = 8.588, p = .000, \eta^2_p = .111$. Both the descriptive and inferential statistics showed that the four outcome measures were not equally difficult but since the mean scores differ only marginally across treatment, time and outcome measure, we attribute no importance to the main of effect of outcome measure and the Time x Outcome Measure interaction in light of the research questions which were investigated.

Table 5.3. Performance on pretest and posttest items by outcome measure and by treatment group
(maximum scores of 8 on every outcome measure and 32 on all pretest or posttest items)

<table>
<thead>
<tr>
<th>Time</th>
<th>Outcome Measure</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>CI95%</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRETEST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GJ</td>
<td>Input practice</td>
<td>36</td>
<td>4.72</td>
<td>1.21</td>
<td>4.31</td>
<td>5.13</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Output practice</td>
<td>35</td>
<td>4.34</td>
<td>1.63</td>
<td>3.78</td>
<td>4.90</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>71</td>
<td>4.54</td>
<td>1.43</td>
<td>4.20</td>
<td>4.87</td>
<td>1</td>
</tr>
<tr>
<td>SR</td>
<td>Input practice</td>
<td>36</td>
<td>5.42</td>
<td>1.48</td>
<td>4.92</td>
<td>5.92</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Output practice</td>
<td>35</td>
<td>5.66</td>
<td>1.53</td>
<td>5.13</td>
<td>6.18</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>71</td>
<td>5.54</td>
<td>1.50</td>
<td>5.18</td>
<td>5.89</td>
<td>3</td>
</tr>
<tr>
<td>CCR</td>
<td>Input practice</td>
<td>36</td>
<td>5.64</td>
<td>1.61</td>
<td>5.10</td>
<td>6.18</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Output practice</td>
<td>35</td>
<td>5.51</td>
<td>1.34</td>
<td>5.06</td>
<td>5.97</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>71</td>
<td>5.58</td>
<td>1.47</td>
<td>5.23</td>
<td>5.93</td>
<td>1</td>
</tr>
<tr>
<td>TR</td>
<td>Input practice</td>
<td>36</td>
<td>4.75</td>
<td>1.46</td>
<td>4.26</td>
<td>5.24</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Output practice</td>
<td>35</td>
<td>4.60</td>
<td>1.19</td>
<td>4.19</td>
<td>5.01</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>71</td>
<td>4.68</td>
<td>1.33</td>
<td>4.36</td>
<td>4.99</td>
<td>2</td>
</tr>
</tbody>
</table>

Note 1. CI = confidence interval; LL = lower limit; UL = upper limit.
Note 2. The within-subjects variable outcome measure was defined as having four levels: (1) grammaticality judgement (GJ), (2) selected response (SR), (3) constrained constructed response (CCR) and (4) translation (TR).
Table 5.3. (Contd.). Performance on pretest and posttest items by outcome measure and by treatment group
(maximum scores of 8 on every outcome measure and 32 on all pretest or posttest items)

<table>
<thead>
<tr>
<th>Time</th>
<th>Outcome Measure</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>CI 95%</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Upper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POSTTEST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GJ</td>
<td>Input practice</td>
<td>36</td>
<td>6.42</td>
<td>1.27</td>
<td>5.99</td>
<td>6.85</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Output practice</td>
<td>35</td>
<td>6.29</td>
<td>1.25</td>
<td>5.86</td>
<td>6.72</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>71</td>
<td>6.35</td>
<td>1.26</td>
<td>6.06</td>
<td>6.65</td>
<td>3</td>
</tr>
<tr>
<td>SR</td>
<td>Input practice</td>
<td>36</td>
<td>6.72</td>
<td>1.19</td>
<td>6.32</td>
<td>7.12</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Output practice</td>
<td>35</td>
<td>6.20</td>
<td>1.64</td>
<td>5.64</td>
<td>6.76</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>71</td>
<td>6.46</td>
<td>1.44</td>
<td>6.12</td>
<td>6.81</td>
<td>2</td>
</tr>
<tr>
<td>CCR</td>
<td>Input practice</td>
<td>36</td>
<td>6.31</td>
<td>1.19</td>
<td>5.90</td>
<td>6.71</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Output practice</td>
<td>35</td>
<td>6.23</td>
<td>1.42</td>
<td>5.74</td>
<td>6.71</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>71</td>
<td>6.27</td>
<td>1.30</td>
<td>5.96</td>
<td>6.57</td>
<td>3</td>
</tr>
<tr>
<td>TR</td>
<td>Input practice</td>
<td>36</td>
<td>5.72</td>
<td>1.91</td>
<td>5.08</td>
<td>6.37</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Output practice</td>
<td>35</td>
<td>5.51</td>
<td>1.48</td>
<td>5.00</td>
<td>6.02</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>71</td>
<td>5.62</td>
<td>1.70</td>
<td>5.22</td>
<td>6.02</td>
<td>1</td>
</tr>
</tbody>
</table>

Note 1. CI = confidence interval; LL = lower limit; UL = upper limit.

Note 2. The within-subjects variable outcome measure was defined as having four levels: (1) grammaticality judgement (GJ), (2) selected response (SR), (3) constrained constructed response (CCR) and (4) translation (TR).
Figure 5.7. Pretest and posttest performance on all four outcome measures by treatment group (maximum scores of 8 with GJ0 = grammaticality judgement pretest, GJ1 = grammaticality judgement posttest, etc.)
Figure 5.8. Pretest and posttest mean performance by treatment group (maximum scores of 8)

Figure 5.9. Interaction between overall mean performance on outcome measures and time of testing (maximum scores of 8)
In light of the results of the two analyses presented above, we did not deem it worthwhile to pursue any subsequent analyses.

5.4 Conclusions and discussion
With respect to the three research questions formulated at the beginning of this chapter, the following conclusions can be drawn. Analyses of the data in Study 2 revealed a large increase from pretest to posttest performance. This increase was a 4.38-point increase when comparing overall pretest and posttest scores (20.32 and 24.70 respectively on a total score of 32). However, it was impossible to state that this effect was mediated by treatment, modality or outcome measure. In other words, the two treatment groups improved their performance from pretest to posttest but that is all that can be said.

We had expected to find a treatment effect. However, anticipating such an effect and considering the limited number of students available as participants, we had decided in advance to refrain once again from including a no-treatment (control) group in Study 2. Since no data from a no-treatment control group are available, it is impossible to say whether the relatively large performance increases obtained were caused by the fact that participants in both groups had received any treatment which dealt with the target structures under investigation.

Although we found additional main effects and interaction effects, these effects were moderate and did not shed any light on the research questions that were drawn up for the study. Consequently, we decided to refrain from interpreting these effects. For a more detailed discussion and interpretation of the findings across all three of the studies, see Chapter 7, Sections 7.2 and 7.3.

5.5 Implications for quantitative data collection in Study 3
In light of the results found after statistical analysis of the quantitative data from Study 2, we decided to make only one fundamental change to the overall research design and research methodology for a third study. The change that was implemented was the inclusion of a no-treatment group. In all other respects, Study 3 was an exact replication of Study 2. The addition of a no-treatment control group was deemed essential for interpreting the data and for drawing conclusions with respect to any possible effect of treatment. Studies 1 and 2 did not have any no-treatment groups so any significant effects that were found in those studies could not be based with any degree of certainty on the treatment that was provided. To allow for more certainty in the interpretation of possible effects, a no-treatment group was added for Study 3. Chapter 6 provides a detailed overview of Study 3.