

Does intensive explicit grammar instruction make all the difference?

Ernesto Macaro and Liz Masterman
University of Oxford, UK

This paper investigates the effect of explicit grammar instruction on grammatical knowledge and writing proficiency in first-year students of French at a UK university. Previous research suggests that explicit grammar instruction results in gains in explicit knowledge and its application in specific grammar-related tasks, but there is less evidence that it results in gains in production tasks. A cohort of 12 students received a course in French grammar immediately prior to their university studies in order to determine whether a short but intensive burst of explicit instruction, a pedagogical approach hitherto unexamined in the literature, was sufficiently powerful to bring about an improvement in their grammatical knowledge and performance in production tasks. Participants were tested at three points over five months, and the results were compared with a group which did not receive the intervention. Our results support previous findings that explicit instruction leads to gains in some aspects of grammar tests but not gains in accuracy in either translation or free composition. Reasons for these findings are discussed in relation to theories of language development and the limitations of working memory.

I Introduction

The research reported in this paper is set within an international debate on the value of an explicit focus on grammar in second language (L2) classrooms (see Doughty and Williams, 1998 for a comprehensive review of the theme). It is also set within a (UK) national modern foreign language (MFL) context which has seen repeated calls, by both theorists and practitioners, for a return to more explicit grammar teaching in schools (Marsden, 1999; Metcalfe *et al.*, 1998; Mitchell, 2000; Wright, 1999), but

Address for correspondence: Ernesto Macaro, Department of Educational Studies, University of Oxford, 15 Norham Gardens, Oxford OX2 6PY, UK; email: ernesto.macaro@edstud.ox.ac.uk

with a distinction made between low-and high-achieving students (McLagan, 1994; Wright, 1999), and for those preparing to study languages at university (Hurman, 1992; Klapper, 1997; 1998; Sheppard, 1993).

What exactly is meant by teaching grammar explicitly is, of course, highly dependent on the viewpoint of the person advocating it or otherwise. For the purposes of this paper, we define it in the way that we believe is meant by many of the authors above:

Establishing as the prime objective of a lesson (or part of a lesson) the explanation of how a morphosyntactic rule or pattern works, with some reference to metalinguistic terminology, and providing examples of this rule in a linguistic, though not necessarily a functional, context.

The research, carried out at the University of Oxford, took place within the context of a mismatch between the university MFL curriculum's emphasis on grammatical awareness and the grammatical knowledge of some of the applicants (see Macaro and Wingate, 2004, for further contextualization). This perceived mismatch may persist even though there was little difference in applicants' overall linguistic performance in external examinations in that a criterion for admission to the university is the achievement of grade A in the nationally standardised Advanced Level MFL examination.

To demonstrate that they can meet this required standard of grammatical knowledge, applicants wishing to study MFL at Oxford must take an admissions grammar test (henceforth AGT¹) in their chosen language(s). The test consists of (*inter alia*) a gap-fill test and translation of isolated sentences into the L2. The intensive grammar course which forms the subject of the present study represented an attempt by the university to bridge this perceived 'grammar gap' for students who had achieved the lowest scores in the AGT (some nine months earlier) but who had nevertheless been accepted on other criteria.

The prime motivation for our study, therefore, was to investigate whether high-achieving students of French who were about to embark on a university course, but had performed relatively badly in an admissions grammar test, might benefit from intensive and explicit instruction in a number of grammatical structures.

With this context in mind, we will now examine two aspects of the research literature relevant to our study. These are: (i) the relationship between explicit and implicit knowledge and (ii) the effectiveness of different approaches to L2 grammar instruction.

II Background

1 *The role of explicit and implicit knowledge*

There has been considerable focus of attention on the relationship between explicit (analysed) grammatical knowledge and implicit (unanalysed) grammatical knowledge and how this might relate to language development. It is generally accepted that explicit knowledge is acquired through controlled processes in declarative memory, while implicit knowledge is acquired through much less conscious or even subconscious processes.

The implications of these two types of knowledge for L2 instruction are twofold. First, if grammar is taught explicitly can it then become automatic so that language can be understood and produced without constant recourse to the rules that generated the explicit knowledge in the first place? Conversely, can language that is acquired implicitly be reflected on if and when the language situation or task demands it? Second, if different language programmes want to measure these different types of knowledge, can they be measured validly?

The degree to which implicitly acquired knowledge is *accessible* has been questioned (Han and Ellis, 1998), for example in production tasks, where knowledge and behaviour are not easily distinguishable. Grammaticality judgement tests (GJT), where subjects have to declare whether a sentence is correct or incorrect, may be more appropriate for distinguishing between knowledge and performance. However, whether such tests really measure implicit (as opposed to explicit) knowledge is now considered to be dependent on a number of conditions, of which the pressure to respond within a given time, and the use of 'rule' rather than 'feel', appear to be the most important (Bialystok, 1979; R. Ellis, 2005).

A difficulty also resides in measuring knowledge about language, in that learners cannot be said to lack explicit knowledge simply because they do not possess the required metalinguistic competence to articulate it. Nevertheless, studies have explored the relationship between metalinguistic knowledge and general language proficiency. For example, Alderson *et al.* (1997) carried out such research among undergraduates and found 'no evidence to support the belief that students with higher levels of metalinguistic knowledge perform better at French, or that they improve their French proficiency at higher rates than other students' (1997: 118) and concluded that 'metalinguistic knowledge and linguistic

proficiency appear to constitute two separate factors of linguistic ability' (1997: 118). This pattern of findings recurs in the study by Han and Ellis (1998), who administered a number of measures, including GJTs and production tasks, to 48 adult learners of English (L2). They claimed that their results lend support to Bialystok's (1982) view that different types of language use draw on different types of knowledge and that metalinguistic knowledge appears unrelated to general language proficiency.

Ellis (2004) has further argued for a distinction between possessing the knowledge itself and the ability to verbalize it, regardless of whether the learner possesses the metalanguage. Does a GJT that also asks the respondent *why* a sentence is incorrect provide evidence of explicit knowledge, or simply an ability to express that knowledge? For example, Green and Hecht (1992) required 300 German learners of English to correct 12 errors in 12 separate sentences and to offer acceptable explanations of the rule. They found that learners were able to correct the errors without being able to offer an explanation of the rules, suggesting implicit knowledge. However, no attempt was made to tap into the learners' thinking processes which may well have provided evidence of controlled reflection on language patterns, rather than implicit knowledge.

These considerations also raise the question as to whether there is a difference between *implicit acquisition* and the *expression of intuitive knowledge*. Grammatical knowledge that is acquired through subconscious processes is not necessarily the same knowledge that is applied intuitively when, for example, one is monitoring one's own production. All kinds of other bits of input and knowledge may be competing for prominence when monitoring intuitively. Ellis (2004: 231) has argued that 'it may be possible to proceduralize explicit knowledge to the point that it cannot be easily distinguished from implicit knowledge'. If this is so, then it may be possible for proceduralized knowledge to be brought back to selective attention for examination because it once was in declarative form. Although these transformations may be true, they do create the situation where the relationship between how knowledge was acquired and how it is demonstrated is not a simple one.

Attention has also centred on the extent to which explicit and implicit knowledge may contribute to variability between a learner's competence and their performance (Ellis, 1985; 1990a; 1990b; Tarone, 1988). In other words it raises the question of how we deal with the discrepancy

between learners' *knowledge about* the L2 and their *ability to use* the L2 in different situations (Bialystok, 1982), leading to an uncertainty regarding how to measure interlanguage development. Macrory and Stone (2000) investigated pre-intermediate learners' acquisition of the perfect tense in French and found not only considerable within-subject variability in the use or non-use of the auxiliary, but also that subjects used an auxiliary (both correctly and incorrectly) in gap-fill exercises while frequently omitting it completely in oral and written production tasks. On a related theme, Hulstijn and de Graaf (1994) theorized that explicit knowledge facilitates the acquisition of implicit knowledge *under certain conditions*. An interesting speculation of theirs, resulting from this line of reasoning, is that explicit instruction may have a greater effect on comprehension tasks than on production tasks.

In sum, although the distinctions between the explicit and implicit acquisition of knowledge, and between intuitive and reflective demonstration of knowledge are still being established, it may nevertheless be the case that the two types of knowledge, although acquired differently, do in fact interact in long term memory. This is sometimes known as the *interface position* (Ellis, 1994; McLaughlin, 1987), a position also held by N. Ellis (2005). It is this interface that continues to interest researchers in that, as a consequence, explicit instruction may indeed have a part to play in developing L2 acquisition.

2 *Is explicit L2 grammar instruction effective?*

The debate regarding whether grammar should be taught explicitly has been a constant one ever since the introduction of the Direct Method in the late-nineteenth century (Richards and Rogers, 1986), which questioned its effectiveness. Despite this questioning, explicit grammar instruction persisted in various forms throughout the twentieth century even though pedagogies of the 1980s and 1990s continued to view explicit grammar teaching with some caution as can be seen from the work of theorists (Prabhu, 1987; Rivers, 1983; Widdowson, 1978); from syllabuses and exam requirements (ILEA, 1979; Edexcel, 2004; AQA, 2005) and from textbooks (Soars and Soars, 1987; Elston and McLagan, 1997). Space does not allow an in-depth analysis of why the debate has persisted, but we can summarize the link between pedagogy and research by identifying two strands of research.

The first strand concerned itself with whether grammatical development was constrained by innate faculties, as in the series of morpheme studies (e.g. Bailey *et al.*, 1974; Larsen-Freeman, 1976) and as in the subsequent interest in teachability (Pienemann, 1984) and in developmental readiness (Spada and Lightbown, 1999; Mackey and Philp, 1998). The second strand concerned itself with the power of input and interaction to deliver acquisition of the rule system without explicit grammar instruction (Krashen, 1985; Long, 1981; 1983; Swain, 1985). Counter to these research strands there has been evidence that, despite immersion in a language, learners continue to make grammatical errors (Harley, 1989), make insufficient progress with competence in low-input courses (Sharwood-Smith, 1981; 1994; Mitchell, 2000), and certain grammatical forms cannot be acquired solely on the basis of comprehensible input (White, 1987). There has also been the notion that academically gifted students might benefit (or at least derive enjoyment) from a teaching style where the focus is on the analysis of the L2 grammar (e.g. Cook, 2001). This argument that academically gifted students may have a special aptitude for linguistic analysis, is one very much put forward by some of the authors cited in the introduction to the context of our study.

Studies of explicit grammar instruction that are relevant to our research fall into two categories: (a) those which have simply compared the relative effectiveness of different instructional approaches on learners' explicit grammatical knowledge, and (b) those which have investigated the relationships between instructional approaches and both explicit grammatical knowledge *and* production. We will start by reviewing studies in the first category.

Scott (1989) taught two university-level classes French relative clauses and the subjunctive using, alternately, an explicit method or an implicit method, and tested them using aural and written gap-filling exercises. The post-tests suggested that both classes made more progress when taught explicitly, but only in the written exercises. However, the author acknowledges that the students had probably been taught previously via an explicit approach and would have been familiar with it. Doughty (1991) investigated the acquisition of English relative clauses under rule-oriented instruction and under meaning-oriented instruction, and whether acquisition of harder ('marked') structures would facilitate acquisition of easier structures. She found that both groups improved

significantly against a control group with no advantage for the rule-oriented instruction. However, instruction in marked relativization did appear to generalize to the acquisition of less marked aspects of the form, suggesting some advantage for focus on form (FonF). Fotos and Ellis (1991) compared explicit instruction in a number of grammatical elements with simply raising university students' consciousness of those elements via communicative pair and group tasks. They concluded, on the basis of GJTs, that consciousness-raising tasks 'appear to be an effective type of classroom activity' (1991: 623) even though they did not necessarily result in the same levels of longer-term learning. In a later investigation, Fotos (1993) found that both instructional approaches were equally effective in promoting subsequent 'noticing' (Schmidt, 1990) of targeted features of the language in written texts and developing longer term knowledge. No measurement of production was administered.

An investigation into more general grammar learning and over a longer period was carried out by Klapper and Rees (2003), who followed a sample of 57 students of German over their four-year degree course. Forty were single- or joint-honours students (Majors) and were taught through an explicit focus on forms (FonFS) approach. The other 17 were studying German 'in combination with a degree in Commerce, Social Science or Law' (2003: 190) and were taught via a FonF approach. We are not told the weighting of the non-language part of the degree in relation to the language part but we are told that both groups received 'almost identical' amounts of language instruction. At the start of the course, there were no significant differences between the two groups, as measured by a C-test and grammar test. By the end of year 2, the FonFS group scored significantly higher in the two types of test. However, at the beginning of year 4, after both groups had studied abroad, the FonF group significantly outperformed the FonFS group in the tests. So, despite the clear advantage of the FonFS group in having received an instructional approach that matched the type of test given, in the long run, the explicit grammar instruction received during the first two years of the course was not the key independent variable.

Studies measuring the impact of instruction on both explicit grammatical knowledge and production have produced conflicting findings. Frantzen (1995) investigated whether explicit grammar teaching and corrective feedback improved grammatical knowledge and accuracy and

fluency of writing, as measured by a discrete-point grammar test and an essay before and after the intervention. Both treatment and comparison groups made significant progress in both areas. However, the experimental group outperformed the comparison group on the grammar test only. A similar study but without a comparison group (Manley and Calk, 1997) found that although some error reduction followed the treatment, there was no holistic improvement in written production.

More positive findings are reported by Leow (1996) who tested undergraduate beginner students of Spanish after 6 hours and after 35 hours of formal exposure to the L2. Significant correlations at around the $R = 0.6$ level were registered between the GJTs and production tasks, suggesting an association between knowledge of the language and performance in it. However, the production tasks were heavily constrained; i.e. students were posed a set of essentially closed questions, rather than being required to generate and monitor their own language.

Using an input-processing approach (VanPatten, 1996; VanPatten and Cadierno, 1993), Benati (2001) investigated the acquisition of the future tense in Italian by three groups of university students. The first group was taught via focus on positive evidence of the inflected form in the input, the second via paradigms to explain the rules followed by output-based practice, and the third, a control, received non-systematic exposure to the target feature. Both the treatment groups outperformed the control group in tests of implicit knowledge, explicit knowledge and oral production. However, in none of the tests did the 'explicit group' outperform the 'input processing group', suggesting no advantage for the explicit explanation of rules. It should be noted, additionally, that the future tense in Italian is a comparatively easy rule.

The line of enquiry regarding easy and hard rules was pursued by Robinson (1996). In a controlled experiment, subjects were taught rules in four different conditions: they viewed sentences and were told it was a memory test; they viewed sentences and were told to look for meaning; they viewed sentences and were told to try to identify rules; they read through the rules that were the focus of the study, then saw some sentences and were asked metalinguistic questions about them. Results showed that simple rules were indeed learnt more easily under all conditions.

Norris and Ortega (2000) concluded, from their meta-analysis of the relative effectiveness of different instructional approaches, that FonF and FonFS are both equally effective and durable. However, they

acknowledged the limitations of conclusions based on a meta-analysis that incorporated vastly different contexts and different outcome measures. They also pointed to variations in the way specific instructional approaches were operationalized, the tendency of explicit treatments to incorporate more instructional components, thereby giving recipients more exposure to the target form than participants in implicit treatment groups, the need for longer post-intervention observation periods to detect significant changes possibly resulting from implicit treatments, and decontextualized, memory-based outcome measures which *de facto* favour the recipients of explicit instruction.

3 Summary

Whilst there is research evidence that some focus on the grammatical features of the L2 is beneficial to developing the interlanguage of a learner, the evidence with regard to the explicit teaching of grammatical features is not sufficiently conclusive to be able to influence pedagogy directly. Particularly inconclusive is the issue of whether being taught rules explicitly leads to successful internalization of those rules. This appears to be linked to the considerable uncertainty over the nature of the relationship between implicit and explicit knowledge, and between implicitly acquired knowledge and the application of knowledge which is, at least in part, intuitive. Nevertheless, the evidence from research tends to support an interaction between these constructs rather than a complete dissociation. For that reason alone, it would seem to be worth continuing to ask questions about explicit grammar instruction. However, there are additional reasons for continuing to investigate the effectiveness of explicit grammar teaching.

First, it remains to be determined whether explicit knowledge can become sufficiently automatic to enable both fluency and accuracy in production tasks. Conversely, we need to explore whether explicit knowledge can be easily brought back under selective attention in order to monitor for possible mistakes in production tasks.

Second, the relationship between implicit and explicit knowledge appears to account for variability between tasks which measure different things. Although GJTs are not without controversy with regard to their validity (Birdsong, 1994; Ellis, 1990a; Gass, 1994), they continue to be used, alongside other types of tasks, in order to test for grammatical

knowledge *of some kind* and in a variety of both research and pedagogical contexts. In this study, a GJT was used in combination with other discrete-point grammar tests to measure knowledge of grammar irrespective as to whether it was intuitive or otherwise and compared this knowledge to that expressed through written production.

Third, evidence from intervention studies suggests that if learners are taught rules explicitly they will perform better in grammar tests. There is tentative evidence that learners will perform better in highly structured production tasks. There is little evidence so far that they will perform better in less structured production tasks. We compared a discrete-point grammar test with two different production tasks: one highly structured and the other less structured.

Fourth, although we have identified many studies involving treatments carried out over a relatively brief period, we were unable to identify any that investigated the effect of a short, but intensive burst of explicit grammar instruction before starting a language course with a strong grammar focus.

Lastly, to our knowledge, no studies tested this kind of instruction on comparatively high-achieving language learners: that is, learners who had been admitted to an 'elite university' through a highly selective admissions process. A related study (Macaro and Wingate, 2004) showed this particular sub-population to be highly motivated to performing well on a grammar-oriented first year of a four-year language programme at that university. In other words, if this particular type of intensive, explicit instruction would be effective with anybody, it would surely be effective with these participants and in these circumstances. Given the national context, other similar institutions may be considering an intensive course of this type. It was therefore appropriate for this type of intervention to be investigated.

III Research questions and rationale

The context outlined above, together with the review of previous research, led us to ask the following inter-related research questions:

- 1) Is an intensive course in explicit French grammar given to high-achieving first-year undergraduates a *sufficiently* powerful intervention to bring about (a) an improvement in their grammatical knowledge, and (b) a reduction in their written production errors?

- 2) Is any immediate improvement sustained over a longer period?
- 3) Is any reduction in production errors brought about without a detrimental effect on other aspects of writing proficiency?
- 4) Is any detected improvement significantly different from the long-term progress made by a comparison group not receiving the intervention?

We laid stress on the concept of *sufficiency* in formulating these research questions as we recognized that other factors beyond our control would almost certainly also influence students' progress, for example, the fact that the students were subsequently taught by different teachers. This variation in learning experience at university is not exclusive to Oxford. Many universities offer additional supported self-study. Moreover, independent self-study, over a period of four to five months, may vary enormously according to a university student's personal motivation. Thus what we would look for would be reliable evidence that the intensive grammar course, 30+ hours of tuition, made a significant difference over and above these factors. If the treatment was 'sufficiently effective', it should result in significant improvement, not only in students' ability to demonstrate knowledge in grammar tasks, but also in controlled and less controlled production tasks. In the less controlled production tasks, we would look to see whether improved accuracy was not obtained at the expense of other aspects of writing proficiency. We formulated research question 2 because of the longitudinal nature of the study with unequal time points of data collection (see below).

IV Method

1 Participants

Participants were drawn from the population of first-year students studying French at the university. The context of the study required us to adopt a purposive strategy by recruiting exclusively those students from a percentile who had achieved the lowest scores in the AGT. The intervention group was recruited by invitation in summer 2003 and numbered 12 students. For the comparison group we identified the tranche of students who had achieved the next lowest scores in the AGT and sent them invitations on their arrival in October 2003. We recruited 10 students who received payment for taking part. All students in the overall sample

Table 1 Summary of AGT scores

Group	N	Lowest	Highest	Mean	SD
Intervention	12	2.4	6.0	5.11	1.01
Comparison	10	4.7	6.6	5.79	0.63

had obtained a grade A (the highest possible) in French in the national examinations at age 18, as required for admission to the university. Table 1 shows the range of AGT scores within the two groups.

With regard to the AGT scores, a Kolmogorov-Smirnov test showed that the distribution of the whole sample was significantly non-normal ($p = 0.013$), and so we could only use non-parametric tests for between-group comparison. However, the distribution within each group was normal, making it possible to conduct parametric tests for repeated measures.

Although the students were not randomly allocated to the two groups, a Mann-Whitney test conducted on their AGT scores showed that the difference between the two groups was not significant ($U = 35.000$, $p = 0.098$). As we shall see, our own pre-tests also did not show any significant difference between the two groups. This, together with their results in national examinations, would suggest that the purposive sampling strategy did not undermine the comparability of the two conditions.

Students in both groups were free to withdraw from the study at any time, but none did so.

2 *The intensive grammar course*

The course tutor, highly thought of and a French native speaker, adopted a FonFS approach, with a preliminary session in which the tutor reinforced students' knowledge of grammatical terminology. The daily programme consisted of three hours of oral and written activities in the classroom (a total of approximately 30 hours), which were designed to develop students' grammatical knowledge by focusing on a list of grammatical topics.

For example, one 60-minute session observed by the second author on day 3 of the course comprised the following sequence of activities, in which the tutor developed the topic of relative pronouns, which he had

begun the previous day, and also revised other grammatical elements which had been covered in earlier sessions:

- 1) Working individually or in pairs, students composed sentences incorporating *que, dont, le quel*, etc. (continuation of previous day's topic) (9 minutes).
- 2) Tutor used a sentence written by a student in a previous exercise to explain the agreement of the direct object in relative clauses with the *passé composé* (3 minutes).
- 3) Tutor explained the use of the relative pronoun *dont* and set an exercise for students to compose sentences. He then coached individual students during the exercise and gave feedback to the whole class (9 minutes).
- 4) Tutor set an exercise for students to work in pairs to reconstruct fragmented sentences involving relative pronouns and the future and conditional verb tenses (15 minutes). Tutor then asked students to read out their work and included explanations of grammatical rules in his feedback. He also asked them to write the sentences in their notebooks (7 minutes).
- 5) Tutor set a computer-based exercise for students to read articles on the Website of *Le Monde*, identify constructions containing relative pronouns and copy them into their notebooks (8 minutes).
- 6) Tutor gave a dictation and asked students to underline occurrences of grammatical elements studied during this session and on the previous day, including demonstrative adjectives and agreement of the past participle (9 minutes).

The daily classroom activities were reinforced by independent study tasks, including translating, essay writing and memorizing verb forms. Students also had twice-weekly sessions, in pairs, with a postgraduate student to help them consolidate what they had learnt.

3 Grammar tests

Three grammar tests were administered to participants during the study as follows:

- *Pre-test*: Intervention group – immediately before the start of the intensive grammar course (September 2003). Comparison group – as soon as practicable after starting their studies (October 2003).

- *Interim test:* Intervention group only – one week after the end of the intensive grammar course and immediately prior to starting their studies (October 2003).
- *Post-test:* Both groups: after 1½ terms' tuition (12 weeks) (February 2004).

The timings enabled us to measure the progress made by the intervention group students, and to gauge the impact of the grammar course plus term-time tuition versus term-time tuition alone.

Although the intensive course covered a number of areas of French grammar, the tests concentrated on four categories only: verbs/tenses/aspect, relative clauses, agreement and prepositions. These had given rise to the greatest proportion of errors, in the first-year exams in June 2003, according to university teaching staff.

The tests were written by the research team as no standardized tests of this type at this level were available. They were checked for accuracy, relevance and consistency, and the 'model answers' for Parts 1–3 were approved by the grammar course tutor. The pre-test was piloted with students studying French in the final year of state secondary education and whose grammatical knowledge could be considered close to that of the sample. They were able to complete the test in the time allowed and gained scores that were considered acceptable.

The tests each lasted 1¼ hours and had an identical four-part structure. The structure and scoring system were as follows:

a Grammaticality judgement (Part 1)

- *Purpose:* Use explicit and/or implicit grammatical knowledge to identify and correct grammatical errors.
- *Structure:* 20 sentences (5 per grammatical category), 12 of which contained a grammatical error.
- *Examples:* **Elle souhaitait qu'il écrirait plus souvent.* (Error in verb tense: *écrivît/écrive*). *La porte par où vous êtes entrés date du XIIIe siècle.* (Testing recognition of use of relative pronoun).
- *Scoring:* Sentences with errors: 1 mark for identifying error + 1 mark for correcting it (max. score: 24). Error-free sentences: 1 mark for leaving sentence unchanged (max. score: 8).

b Error correction and rule explanation (Part 2)

- *Purpose:* Show evidence of grammatical knowledge by correcting grammatical errors and show evidence that grammatical rules taught have been retained.
- *Structure:* 5 sentences, each containing a grammatical error.
- *Example:* **Mon oncle a été mordu d'un énorme tigre blanc.* (Error: preposition should be *par*, not *de*. Rule: 'The agent of a passive verb describing an action or event is always preceded by the preposition *par*.')
- *Scoring:* Error correction component: up to 2 marks for each error correction. Rule explanation: 1 mark for correctly stating (or paraphrasing) the grammatical rule (regardless of whether grammatical terminology is used) + 1 mark for the use of some appropriate grammatical terminology in that explanation. (Max. score: 20.)

c Translation (Part 3)

- *Purpose:* Test accuracy in students' productive use of French in a situation of relatively controlled output.
- *Structure:* Five sentences, which together contained a total of three instances of each of the four grammatical categories being tested.
- *Examples:* *The people I met in France during the summer will be coming to England for Christmas. Les gens que j'ai rencontrés [connus] en France pendant l'été vont venir [viendront] en Angleterre pour [à] Noël.* (Verb tense: *vont venir [viendront]*. Agreement: *rencontrés [connus]*. Preposition: *en* (two occurrences), *pour [à]*. Relative pronoun: *que*.)
- *Scoring:* 1 mark for each correctly translated structure (remainder of each sentence was disregarded). Incorrect vocabulary was admissible if the gender was correct. (Max. score: 12.)

d Narrative composition (Part 4)

- *Purpose:* Test students' accuracy and general writing proficiency in a production task in which they were relatively free to form their own output.
- *Structure:* Re-tell a story sequence of six pictures as a third-person narrative in the past tense, in at least 250 words.

- *Scoring:* (a) Holistic scores: accuracy, lexical diversity, range of expression, using a nationally standardized marking scheme (Oxford Delegacy of Examinations) with a maximum score of 10 each; (b) Computed scores: accuracy (percentage of correct noun and verb phrases); lexical diversity using Measure D (Malvern *et al.*, 2004).

Part 1 and the error-correction component of Part 2 were scored by the second author in accordance with the model answers approved by the course tutor. Students' rule explanations in Part 2 were open to subjective interpretation and were therefore scored by both authors, with an average inter-rater reliability of 75%. Agreement was then reached over the remaining items.

In the narrative task (Part 4), when scoring the compositions holistically, the two authors marked the compositions separately, then reached agreement on the scores by discussion. The computational methods were intended to triangulate these scores and to counteract subjective elements in the holistic marking. We were unable, however, to identify a satisfactory *computational* measure for range of expression.

All scores were entered on SPSS version 11.

V Results

1 Discrete-point grammar tests within groups

In order to begin answering research questions 1 and 2, we first analysed the within-group change over time. The discrete-point grammar tests (Parts 1 and 2) were computed for frequencies and their mean scores compared at each time point using a paired samples t-test. These are summarized in Table 2. Part 1 in the table is split in two, to reflect the different marking schemes used with erroneous and error-free sentences.

The mean ability of the intervention group to spot and correct errors, and to provide explanations for their corrections, improved significantly over the period of the tests. However, it should be noted that the ability to spot errors improved significantly between interim- and post-test rather than between pre- and interim-test. Moreover, the improvement in their ability to recognize error-free sentences was less marked, and even declined between the interim and post-tests. It should also be noted

Table 2 Within-groups mean scores (and standard deviations) for parts 1 and 2 of grammar tests.

	Cond'n	Pre-	Interim	Post-	Pre- to Interim	Interim- to Post-	Pre- to Post-
1a. Erroneous sentences (max. 24)	I	5.00 (3.84)	7.58 (5.20)	12.54 (3.90)	t = 1.705 p = 0.116	t = 3.224 p = 0.008	t = 5.387 p = 0.000
	C	7.90 (0.74)		9.15 (4.26)			t = 1.168 p = 0.273
1b. Error-free sentences (max. 8)	I	2.33 (1.37)	4.08 (1.78)	3.17 (1.70)	t = 2.836 p = 0.016	t = 3.224 p = 0.008	t = 1.820 p = 0.096
	C	2.90 (2.59)		3.00 (1.49)			t = 0.218 p = 0.832
1c. Combined score (max. 32)	I	7.33 (3.89)	11.67 (5.31)	15.71 (4.15)	t = 2.422 p = 0.034	t = 2.226 p = 0.048	t = 6.204 p = 0.000
	C	10.80 (4.18)		12.15 (4.20)			t = 1.125 p = 0.290
2. Error correction & rule expl'n (max. 20)	I	9.13 (3.21)	9.29 (3.96)	12.63 (4.21)	t = 0.133 p = 0.897	t = 1.984 p = 0.073	t = 2.440 p = 0.033
	C	10.75 (2.06)		8.20 (3.29)			t = 1.980 p = 0.079

Notes: Boldface indicates a significant difference.

I = Intervention condition; C = Comparison condition.

that the intervention group's ability to correct errors and provide rule explanations improved significantly only between pre- and post-test, and not between pre- and interim test.

In contrast, the comparison group's scores did not improve significantly and, in the error correction and rule explanation task, declined. Standard deviations indicate some wide variations in individual scores and this issue will be address later.

2 Production tasks within groups

Table 3 summarizes the results of the translation.

Neither group improved significantly between pre- and post-test, and there was even a slight decline in the intervention group's mean performance between interim and post-test. Analysis of individual scripts confirmed that students rarely used avoidance strategies (i.e. simplifying the text to be translated). This was detected in only 15 out of 672 cases

Table 3 Within-groups mean scores (and standard deviations) for translation (part 3)

	Cond'n	Pre-	Interim	Post-	Pre- to Interim	Interim- to Post-	Pre- to Post-
3. Translation (max.12)							
	I	6.58 (1.72)	7.12 (1.90)	6.92 (2.19)	t = 1.096 p = 0.297	t = 0.405 p = 0.693	t = 0.471 p = 0.647
	C	7.35 (1.73)		7.80 (1.75)			t = 0.927 p = 0.378

(i.e. 2.23%). In other words, the translation was indeed measuring controlled output.

Table 4 summarizes both the holistic and the computed scores for all the assessments performed on students' narrative compositions, showing that, overall, both groups' performance improved significantly over the period of the study, and with generally smaller variations than in the preceding three tasks. It should be noted that the intervention group did not improve on grammatical accuracy (either measured holistically or computed) between pre- and interim tests.

In terms of within-groups differences, the computational scores produced similar scores to the holistic marking, with the exception of the intervention group's accuracy, in which the computation detected a slight dip from pre- to interim test, and the comparison group's lexical diversity, which the Measure D calculation does not show as having expanded significantly.

3 *Between-group comparisons*

In order to further answer research questions 1, 2 and 3, and specifically answer question 4, we carried out between group comparisons. Mann-Whitney calculations were performed for the reasons given above. Table 5 shows the results for each part of the tests. The second and fourth columns show which group obtained the higher mean score for the relevant part (for the actual figures, please refer to the preceding tables).

In the pre-test, none of the differences were significant between the two conditions, reflecting the absence of a significant difference in the AGT scores of the two groups. In the post-test, there was a significant

Table 4 Within-groups mean scores (and standard deviations) for narrative composition (part 4)

	Condn	Pre-	Interim	Post-	Pre- to Interim	Interim- to Post-	Pre- to Post-
4a. Holistic: accuracy (max. 10)	I	4.93 (1.56)	5.50 (0.80)	6.50 (0.67)	t = 1.865 p = 0.089	t = 5.745 p = 0.000	t = 4.710 p = 0.001
	C	4.70 (1.06)		6.80 (0.79)			t = 6.034 p = 0.000
4b. Holistic: lexical diversity (max. 10)	I	5.58 (0.90)	6.50 (0.67)	6.92 (0.67)	t = 3.527 p = 0.005	t = 1.449 p = 0.175	t = 4.690 p = 0.001
	C	6.00	(1.25)	7.40 (0.84)			t = 3.500 p = 0.007
4c. Holistic: range of exprn (max. 10)	I	5.17 (1.27)	6.08 (0.79)	6.75 (0.62)	t = 2.727 p = 0.020	t = 2.602 p = 0.025	t = 5.062 p = 0.000
	C	5.14 (1.10)		6.70 (0.68)			t = 3.748 p = 0.005
4d. Computed: accuracy (max. 100%)	I	66.58 (9.38)	65.72 (7.51)	73.12 (8.33)	t = 0.398 p = 0.698	t = 3.674 p = 0.004	t = 3.401 p = 0.006
	C	64.64 (5.85)		73.96 (6.11)			t = 3.821 p = 0.004
4e. Computed: lexical diversity: (no max.)	I	70.45 (12.45)	85.25 (12.18)	90.40 (16.89)	t = 3.157 p = 0.009	t = 0.791 p = 0.446	t = 4.620 p = 0.001
	C	73.84 (11.20)		83.70 (18.70)			t = 1.794 p = 0.106

Notes: Boldface indicates a significant difference.

I = Intervention condition; C = Comparison condition.

between-groups difference in *one score only*, the combined error correction and rule explanation task, and this may be accounted for in part by an actual decline in the comparison group's mean performance (see Table 1).

To check whether these results were distorted by the small sample sizes, we calculated the effect size – Cohen's *d* – for each task in the post-test by dividing the differences between the means of the two groups by the pooled standard deviation of both groups, as advised by Norris and Ortega (2000). A *d* value approaching 1 is deemed large (i.e. the findings carry weight). In our study, this was the case only with the 'erroneous sentence' component of Part 1 (*d* = 0.835), where the Mann-Whitney test approached significance, and Part 2 (*d* = 1.167), although the latter figure needs to be treated with caution

Table 5 Statistical analysis of the differences between the two groups in their pre- and post-test scores

	Pre-test: higher mean score	Pre-test: between-groups difference	Post-test: higher mean score	Post-test: between-groups difference
1a. Erroneous sentences	C	U = 38.500 p = 0.159	I	U = 31.500 p = 0.059
1b. Error-free sentences	C	U = 45.500 p = 0.346	I	U = 56.500 p = 0.821
1c. Combined score (1a + 1b)	C	U = 34.500 p = 0.093	I	U = 33.000 p = 0.080
2. Error correction & rule explanation	C	U = 40.500 p = 0.203	I	U = 25.500 p = 0.021
3. Translation	C	U = 44.00 p = 0.314	C	U = 52.50 p = 0.628
4a. Holistic: accuracy	I	U = 57.50 p = 0.872	C	U = 47.00 p = 0.418
4b. Holistic: lexical diversity	C	U = 51.00 p = 0.582	C	U = 41.00 p = 0.228
4c. Holistic: range of expression	I	U = 57.00 p = 0.872	I	U = 57.00 p = 0.872
4d. Holistic: total	C	U = 59.00 p = 0.974	C	U = 48.00 p = 0.628
4e. Computed: accuracy	C	U = 57.00 p = 0.872	I	U = 52.00 p = 0.628
4f. Computed: lexical diversity	C	U = 46.00 p = 0.381	I	U = 45.00 p = 0.346

Note: Boldface indicates a significant difference.

as it lies just outside upper bound (1.13) of the 95% confidence level computed by Norris and Ortega in their meta-analysis of metalinguistic tasks. In other words, sample size did not appear to influence results.

4 The intensive grammar course: a sufficiently powerful intervention?

On the basis of the results reported so far, any impact of the intensive grammar course appears to have been limited to an improvement in students' ability to identify and correct ungrammatical sentences and to identify and explain errors when sensitized to their presence (as shown

by scores from Parts 1 and 2 of the tests). In our research questions, however, we drew attention to the concept of sufficiency and to the notion of sustained improvement.

To shed light on these, we asked the whole sample, after the post-test (i.e. four months into their studies), whether they had attended weekly grammar classes organized centrally by the University in addition to the language tuition which they received. Thirteen said that they had. We then mapped students' total scores for *Parts 1 and 2 only* in each test to their response to this question, as shown in Figure 1.

A feature of Figure 1 is the contrast between the general pre- to post-test improvement shown by the intervention group, and the mixed performance by the comparison group students, a substantial number of

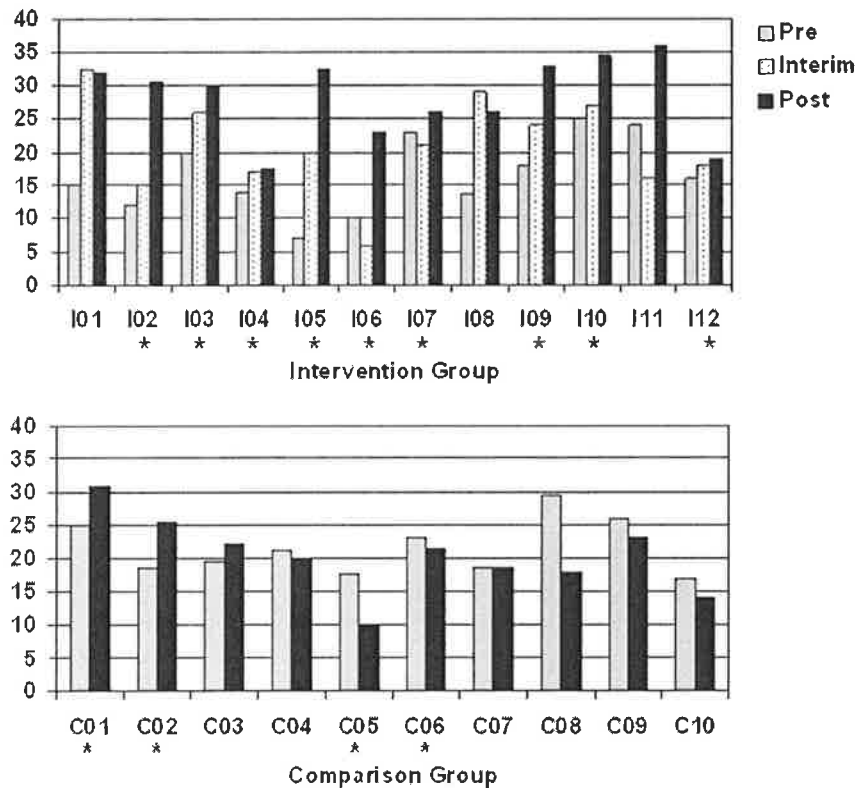


Figure 1 Students' total scores for parts 1 and 2 in the tests (grammaticality judgement and error correction/rule explanation). Asterisks denote students who attended additional grammar classes.

whom actually performed worse on post-test after four months of university instruction.

With the exception of student I05, the students in the intervention group appear to fall into two roughly equal categories in relation to their improvement in the discrete-point grammar tasks: those who showed the most improvement immediately after the grammar course, i.e. at interim test (I01, I08, I03, I04, I12) and those who showed little immediate improvement (or even an initial decline) but subsequently leaped forward (I02, I06, I09, I11, I10, I07). For students in the second category, either the course had a delayed effect or its benefit was combined with subsequent experience of teaching and study. Effect size calculations using the intervention group's pre-, interim and post-test scores from Figure 1 were even more explicit, with Cohen's *d* calculated as 0.696 (i.e. moderate to strong) for the contrast between pre- and interim tests and 1.109 (i.e. very strong) for the contrast between interim and post-tests.

It would appear that many of the 'most-improved' students were also attending the extra grammar classes. Indeed a significant correlation was discovered between the claimed attendance and the percentage improvement in students' test scores between October 2003 and February 2004 (i.e. interim to post-test for the intervention group, pre- to post-test for the comparison group): Spearman's $\rho = 0.503$, $p = 0.017$, $N = 22$.

VI Discussion

The study described in this paper set out to answer four inter-related questions regarding whether an intensive course in French grammar given to high-achieving first-year undergraduates, prior to starting their degree programme proper, was a *sufficiently* powerful intervention to bring about an improvement in their grammatical knowledge, both in the short and long term, a reduction in their production errors without any detrimental effect on other aspects of written production, and these improvements as compared to a group who did not receive the intensive course.

Our findings suggest that the intensive grammar course was *not* a sufficient factor to bring about a significant improvement in their grammatical knowledge as there was no greater ability to make judgements *overall* of grammaticality when compared to the comparison group. There was, however, evidence of a greater ability to correct errors in

sentences once these had been judged ungrammatical. Whether students were applying explicit knowledge gained from the course, or whether they were correcting 'by feel' cannot be ascertained, as there was no apparent parallel improvement in their ability to articulate rules.

The intensive grammar course did not bring about a reduction in production errors in either controlled or uncontrolled production tasks. There is no evidence that explicit grammar taught on the course led to effective production monitoring. Our earlier review would suggest that this production monitoring could have resulted from one or more of three possible factors: (i) the proceduralization of explicit knowledge so that certain elements would now 'feel' incorrect; (ii) the 'bringing back' of proceduralized knowledge under selective attention and the consequent application of a particular rule; (iii) the direct application of explicit knowledge which had never become proceduralized. Interestingly, the intervention group did not outperform the comparison group in the translation task, a controlled output task that tested the same range of grammatical structures as the GJT.

With specific regard to research question 2, we investigated whether any immediate improvement resulting from the intervention was sustained over the longer period. In the majority of measures there was no significant improvement in the intervention group's performance on the interim-test. We have noted above that the intervention group was divided roughly into those who made some progress immediately after the course and those whose progress was delayed. For those students who demonstrated an immediate improvement after the intervention, this was not sustained over the longer period. Since the interim test was taken only one week after the end of the grammar course, it is clear that the short-term impact of explicit instruction was minimal and, in some cases, negative. On the other hand some students did appear to make progress between interim and post-test but we attribute this to a possible awareness raising brought about by the intervention, which was then consolidated by later grammar study. In the absence of other data, we can only speculate that individual differences may be the cause of the different effects of the intervention. Developmental readiness for the acquisition of certain elements of grammar may have been spurred by the intervention and consolidated only if those elements were reinforced through attendance at subsequent grammar classes or through self-study with a grammar focus. If true, this speculation would lend support to

previous research on developmental readiness, although a qualitative analysis through case studies of each student's interlanguage development would be necessary to provide the desired insights.

Our findings, then, provide tentative support to studies previously cited (such as Green and Hecht, 1992) that there may be some link between explicit learning and its application in discrete-point grammar tasks. However, this evidence appears to hold for only some students rather than being a relative but general trend in all students. For some students it is possible that a burst of intensive exposure to a large number of grammatical forms and rules may actually have served to confuse them in the short term.

Our findings also lend support to the conclusions of previous studies (e.g. Bialystok, 1982) that linguistic knowledge manifests itself differently according to task requirements. We should note that in our case this applies to both the highly structured translation task and to the freer composition. The question is why?

One possible explanation may reside in the limitations associated with selective attention and working memory (Baddeley, 1986; 1997) and in the effect of multi-tasking (Skehan, 1998). When attempting to identify incorrect sentences, the subject's cognitive processes are almost entirely focused on identifying errors, while in production tasks the attention is distributed over a number of problems to be solved almost simultaneously.

A second explanation may be related to variability in interlanguage (Ellis, 1985) and to retrieval processes. Allowing for the co-existence of more than one form for a particular element in long-term memory, different psychological mechanisms may be involved (a) in processing an element produced by others and perceived through the visual medium, and (b) in processing an element retrieved and then produced by the self, monitored (at least in the first instance) through the phonic medium for correctness. In other words, a receptive skill such as reading may allow for more direct trace connections to the stronger of the systematic variables than do production skills. The stronger of the systematic variables may well have been influenced by visually displayed positive input. We should also note that, when reading, the learner decodes first via visual connections and then, if necessary, via phonic connections (especially in a GJT). In contrast, when rehearsing for production a subject *may* be encoding via phonic connections before visual connections. Given the

lack of orthographic transparency of the French language, this feature may be particularly accentuated by language specificity. A related explanation might be that, when reading, a learner is being exposed directly to sentences and asked to judge their correctness, while in production they may first be generating phrases or utterances which compete with subsequent sentences being monitored (either in rehearsal or on the page). These observations support the earlier conclusions by Hulstijn and de Graaf (1994) cited above. Further research may illuminate these hypotheses.

Our prime concern in measuring range of expression and lexical diversity was to ensure that other aspects of the writing process were not adversely affected by explicit grammar tuition in case the post-test results proved more positive for grammatical accuracy in free production (research question 3). Since this was not so, these data are largely redundant for the purposes of this paper.

VII Limitations

Our study, of course, was not without its limitations. These were imposed largely by the non-random nature of the selection and the small number of participants. Both of these factors lay outside the researchers' control due to the 'remedial purpose' of the grammar course and the reliance on volunteers for participation in the comparison group. It was thus a matter of fortune rather than of deliberate design that the differences in the AGT scores between the two groups proved statistically non-significant, as did the pre-tests. Even though effect size calculations reinforced the significance of our principal findings, our lack of control over other variables – such as the quantity and quality of regular language teaching received by students and individual learner differences in approaches to self-study – means that we cannot generalize the results with confidence beyond our sampling frame of Oxford MFL students.

VIII Conclusions

The intensive course of explicit grammar teaching was not a sufficiently powerful independent variable in bringing about the intended structural change in the intervention group's interlanguage. In other words, it did *not* 'make all the difference'.

There are a number of implications that we feel we can draw despite the limitations above.

First, this type of course may be inappropriate – and even ineffective – for university courses where the prime objective is students' development of the four language skills. However, it might prove useful for university courses which emphasize what we might call an *awareness* of the structures of language, if only to boost the confidence of students who have not received this kind of explicit tuition previously.

Second, our findings do not lend support to the argument that it is the linguistically most able students who are most likely to benefit from intensive and explicit grammar instruction. Since exposure to language analysis in an intensive course such as this did not result in a great improvement in grammatical awareness, and none in production, the implication is that this sub-population of students should not be treated differently unless for course-specific purposes.

Third, since our findings support previous evidence which has suggested that the development of grammatical accuracy (i) cannot easily be hurried, (ii) is individually developed, and (iii) requires continuous exposure to both positive and negative evidence in both receptive and productive tasks, then the 'short, sharp shock' of an intensive two-week course does not match these developmental patterns. However, this does not necessarily mean that *in combination* with other pedagogical interventions it might not usefully contribute to development. For example, were it to be associated with subsequent *sustained* FonF instruction, it might indeed prove beneficial. On the basis of our findings, we hope to be able to investigate this possibility in the future.

Acknowledgements

We acknowledge funding for both the grammar course and the associated research study from the Higher Education Funding Council for England (HEFCE). We also thank the following for their contribution to the study: Michaël Abecassis, Kate Tunstall, Ursula Wingate, Suzanne Graham, Brian Richards, Robert Vanderplank and Lynn Erler.

We would also like to acknowledge the two anonymous reviewers for their constructive comments in helping to shape this article.

Note

- ¹ It should be made clear that neither the authors nor the university consider the AGT a standardized test of grammatical knowledge. It is merely a component in the admissions process designed to provide an indicator of current grammatical knowledge.

IX References

- Alderson, J.C., Clapham, C. and Steel, D.** 1997: Metalinguistic knowledge, language aptitude, and language proficiency. *Language Teaching Research* 1(2): 93–121.
- AQA** 2005: Qualifications and subjects. Retrieved 24 February 2005 from <http://www.aqa.org.uk/qual/>.
- Baddeley, A.** 1986: *Working memory*. Clarendon Press.
- 1997: *Human memory: theory and practice*. Psychology Press.
- Bailey, N., Madden, C. and Krashen, S.** 1974: Is there a ‘natural sequence’ in adult second language learning? *Language Learning* 21: 235–43.
- Bautier-Castaing, E.** 1977: Acquisition comparée de la syntaxe du Français par des enfants francophones et non francophones. *Études de linguistique appliquée* 27: 19–41.
- Benati, A.** 2001: A comparative study of the effects of processing instruction and output-based instruction on the acquisition of the Italian future tense. *Language Teaching Research* 5(2): 95–127.
- Bialystok, E.** 1979: Explicit and implicit judgements of L2 grammaticality. *Language Learning* 29(1): 81–103.
- 1982: On the relationship between knowing and using forms. *Applied Linguistics* 3: 181–206.
- Birdsong, D.** 1994: Decision making in second language acquisition. *Studies in Second Language Acquisition* 16(2): 169–82.
- Brown, R.** 1973: *A first language: the early stages*. Harvard University Press.
- Cook, V.** 2001: *Second language learning and language teaching*, third edition. Arnold.
- Doughty, C.** 1991: Second language instruction does make a difference. *Studies in Second Language Acquisition* 13(4): 431–69.
- Doughty, C. and Williams, J.**, editors, 1998: *Focus on form in classroom second language acquisition*. Cambridge University Press.
- Dulay, H. and Burt, M.** 1974: Natural sequences in child second language acquisition. *Language Learning* 24: 253–78.
- Edexcel** 2004: Edexcel qualifications. Retrieved 24 February 2005 from <http://www.edexcel.org.uk/qualifications/>.

- Ellis, N.** 2005: At the interface: dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition* 27(2): 305–52.
- Ellis, R.** 1985: Sources of variability in interlanguage. *Applied Linguistics* 6(2): 118–31.
- 1990a: Grammaticality judgements and learner variability. In Burmeister, H. and Rounds, P.L., editors, *Variability in second language acquisition: Proceedings of the Tenth Meeting of the Second Language Research Forum*. University of Oregon, Department of Linguistics, 25–60.
- 1990b: A response to Gregg. *Applied Linguistics* 11: 384–91.
- 1991: Grammaticality judgements and second language acquisition. *Studies in Second Language Acquisition* 13: 161–86.
- 1994: *The study of second language acquisition*. Oxford University Press.
- 2004: The definition and measurement of L2 explicit knowledge. *Language Learning* 54(2): 227–75.
- 2005: Measuring implicit and explicit knowledge of a second language: a psychometric study. *Studies in Second Language Acquisition* 27(2): 141–72.
- Elston, T. and McLagan, P.** 1997: *Génial*. Oxford University Press.
- Fotos, S.** 1993: Consciousness-raising and noticing through focus on form: grammar task performance vs. formal instruction. *Applied Linguistics* 14(4): 385–407.
- Fotos, S. and Ellis, R.** 1991: Communicating about grammar: a task-based approach. *TESOL Quarterly* 25: 605–28.
- Frantzen, D.** 1995: The effects of grammar supplementation on written accuracy in an intermediate Spanish content course. *Modern Language Journal* 79(3): 329–55.
- Gass, S.** 1994: The reliability of second language grammaticality judgements. In Tarone, E., Gass, S. and Cohen, A.D., editors, *Research methodology in second language acquisition*. LEA, 303–22.
- Green, P. and Hecht, K.-H.** 1992: Implicit and explicit grammar: an empirical study. *Applied Linguistics* 13(4): 168–84.
- Han, Y. and Ellis, R.** 1998: Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research* 2: 1–23.
- Harley, B.** 1989: Functional grammar in French immersion: a classroom experiment. *Applied Linguistics* 10: 331–59.
- Hulstijn, J. and de Graaff, R.** 1994: Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? *AILA Review* 11: 97–112.

- Hurman, J.** 1992: Performance in the A level speaking test by candidates with GCSE training: oral examiners' views. *Language Learning Journal* 5: 8–10.
- ILEA** 1979: *Graded objectives for modern foreign languages*. HMSO.
- Klapper, J.** 1997: Language learning at school and university: the great grammar debate continues (I). *Language Learning Journal* 16: 22–27.
- 1998: Language learning at school and university: the great grammar debate continues (II). *Language Learning Journal* 18: 22–29.
- Klapper, J.** and **Rees, J.** 2003: Reviewing the case for explicit grammar instruction in the university foreign language learning context. *Language Teaching Research* 7(3): 285–314.
- Krashen, S.** 1982: *Principles and practice in second language learning and acquisition*. Pergamon.
- Krashen, S. D.** 1985: *The input hypothesis: issues and implications*. Longman.
- Larsen-Freeman, D.** 1976: An exploration of the morpheme order of second language learners. *Language Learning* 26: 125–34.
- Leow, R.** 1996: Grammaticality judgements tasks and second-language development. In Alatis, G., Straehle, C., Rankin, M. and Gallenberger, B., editors, *Georgetown University round table on language and linguistics 1996*. Georgetown University Press, 126–39.
- Long, M.** 1981: Input, interaction and second language acquisition. In Winitz, H., editor, *Native language and foreign language acquisition. Annals of the New York Academy of Sciences*, Vol. 379. New York Academy of Sciences, 259–78.
- 1983: Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics* 4: 126–41.
- Macaro, E.** and **Wingate, U.** 2004: From sixth form to university: motivation and transition among high achieving state-school language students. *Oxford Review of Education* 30(4): 467–89.
- Mackey, A.** and **Philp, J.** 1998: Conversational interaction and second language development: recasts, responses and red-herrings? *Modern Language Journal* 82: 283–324.
- McLagan, P.** 1994: Grammar and the 'less able' learner. In King, L. and Boaks, P., editors, *Grammar! A conference report*. CILT, 69–73.
- McLaughlin, B.** 1987: *Theories of second language learning*. Edward Arnold.
- Macrory, G.** and **Stone, V.** 2000: Pupil progress in the acquisition of the perfect tense in French: the relationship between knowledge and use. *Language Teaching Research* 4(1): 55–82.

- Malvern, D., Richards, B., Chipere, N. and Durán, P.** 2004: *Lexical diversity and language development: quantification and assessment*. Palgrave.
- Manley, J. and Calk, L.** 1997: Grammar instruction for writing skills: do students perceive grammar as useful? *Foreign Language Annals* 39: 73–83.
- Marsden, R.** 1999: Go for the ‘a-ha!’ factor in grammar learning. *Deutsch: Lehren und Lernen* 19: 15–16.
- Metcalfe, P., Laurillard, D. and Mason, R.** 1998: ‘It’s just a word’: pupils’ perceptions of verb form and function. *Language Learning Journal* 17: 14–20.
- Mitchell, R.** 2000: Applied linguistics and evidence-based classroom practice: the case of foreign language grammar pedagogy. *Applied Linguistics* 21(3): 281–303.
- Norris, J.M., and Ortega, L.** 2000: Effectiveness of L2 instruction: a research synthesis and qualitative meta-analysis. *Language Learning* 50: 417–528.
- Pienemann, M.** 1984: Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition* 6(2): 186–214.
- Prabhu, N.S.** 1987: *Second language pedagogy*. Oxford University Press.
- Richards, J.C. and Rogers, T.S.,** editors, 1986: *Approaches and methods in language teaching*. Cambridge University Press.
- Rivers, W.M.** 1983: *Communicating naturally in a second language: theory and practice in language teaching*. Cambridge University Press.
- Robinson, P.** 1996: Learning simple and complex second language rules under implicit, incidental, rule-search and instructed conditions. *Studies in Second Language Acquisition* 18(1): 27–68.
- Schmidt, R.** 1990: The role of consciousness in second language learning. *Applied Linguistics* 11: 129–58.
- Scott, V.** 1989: An empirical study of explicit and implicit teaching strategies in French. *Modern Language Journal* 72: 14–22.
- Sharwood-Smith, M.** 1981: Consciousness raising and the second language learner. *Applied Linguistics* 2: 159–69.
- 1994: *Second language learning: theoretical foundations*. Longman.
- Sheppard, R.** 1993: Getting down to brass syntax: German teaching and the great standards debate. *German Teaching* 8: 3–9.
- Skehan, P.** 1998: *A cognitive approach to language learning*. Oxford University Press.
- Soars, J. and Soars, L.** 1987: *Headway*. Oxford University Press.
- Spada, N. and Lightbown, P.M.** 1999: Instruction, first language influence, and developmental readiness in second language acquisition. *Modern Language Journal* 83(1): 1–22.

- Swain, M.** 1985: Communicative competence: some roles of comprehensible input and comprehensible output in its development. In Gass, S. and Madden, C., editors, *Input and second language acquisition*. Random House, 235–53.
- Tarone, E.** 1988: *Variation in interlanguage*. Edward Arnold.
- VanPatten, B.** 1996: *Input processing and grammar instruction*. Ablex.
- VanPatten, B.** and **Cadierno, T.** 1993: Explicit instruction and input processing. *Studies in Second Language Acquisition* 15: 225–43.
- White, L.** 1987: Against comprehensible input: the input hypothesis and the development of second language competence. *Applied Linguistics* 8: 95–110.
- Widdowson, H.G.** 1978: *Teaching language as communication*. Oxford University Press.
- Wright, M.** 1999: Grammar in the languages classroom: findings from research. *Language Learning Journal* 19: 33–39.

Copyright of Language Teaching Research is the property of Arnold Publishers and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.