

UNIVERSITY OF SOUTHAMPTON

FACULTY OF HUMANITIES, ARTS & SOCIAL SCIENCES

School of Education

*Teaching and Learning of French Verb Inflections: A Classroom Experiment Using
Processing Instruction*

by

Emma Josephine Marsden

Thesis for the degree of Doctor of Philosophy

March 2004

UNIVERSITY OF SOUTHAMPTON
ABSTRACT
FACULTY OF HUMANITIES, ARTS & SOCIAL SCIENCES
SCHOOL OF EDUCATION
Doctor of Philosophy
TEACHING AND LEARNING OF FRENCH VERB INFLECTIONS: A CLASSROOM
EXPERIMENT USING PROCESSING INSTRUCTION
by Emma Josephine Marsden

This study contributes to 'Focus on Form' research, investigating effective ways of teaching grammar within a broadly communicative Modern Foreign Languages curriculum. Currently in the UK there is a widely felt dissatisfaction that many school learners' ability to create their own language in target-like ways remains problematic, in particular for the development of the highly inflected French verb system. Sustained work suggests that 'Processing Instruction' (PI) can promote the learning of certain language features. However, no study has yet explored one of the principles behind PI (that learners do not make learning gains if given input activities that allow a more incidental/implicit mode of processing, as learners do not tend to interpret the meaning of items of low communicative value).

This study was conducted in two UK schools / year 9 classrooms. It adopted a quasi-experimental design to compare PI with Enriched Input (EI), which focussed learners' attention on lexical items and / or overall sentential meaning, but allowed an implicit mode of learning the target verb inflections. This shares many similarities with current listening and reading activities in UK classrooms. Progress in learning the target features was also monitored in a non-active control class in school 1. A battery of pre, post and delayed post tests was used to assess the short and longer term impacts on learners' ability to understand and use, in oral and written modalities, a selection of French verb inflections in the present and perfect tenses. Lessons were monitored in all three classes prior to and throughout the study and the pupils' and teachers' reactions to the materials were surveyed to strengthen the study's validity.

In general, instructional type *alone* did not have a significant impact on the pupils' learning. However, it was found that in school 2 (class B) the learners who experienced PI made and maintained statistically significant learning gains in all the measures taken, whereas the EI learners did not. This suggests there are potential benefits of using PI to promote the learning of verb inflectional system with such learners.

In contrast, in school 1 (Class A), both the EI and PI learners made and maintained statistically significant learning gains, suggesting that with these learners a more incidental mode of processing was equally beneficial. This may be due to a background school ethos of teaching and testing grammar, as the parallel, non-active control class also made some gains between pre and post test. However, this does not account for the *extent* of the gains by EI learners in class A in the listening, reading and writing measures. It is therefore suggested that the EI learners in class A, unlike those in class B who were at a lower developmental stage, had sufficient processing resources to interpret the meaning of verb inflections as well as lexical items and/or sentential meaning (i.e. engage in incidental / multiple processing). In addition, it is acknowledged that it is possible that more general characteristics of the normal class teaching in class A (e.g. sequencing of grammar pedagogy tasks) may have enabled learners in class A to benefit from EI more than learners in class B.

The study recommends laboratory based research to explore further the existence / nature of incidental input processing amongst learners of different developmental stages. This study also highlights how contextual information can shed essential light on findings in educational experimental research.

Table of Contents

Lists of figures & tables	i
Author's declaration	iv
Acknowledgements	v
INTRODUCTION: OVERVIEW OF THE THESIS	1
CHAPTER 1 RATIONALE	2
1.1 Dissatisfaction with the teaching and learning of 'language form'	2
1.2 Current rehabilitation of grammar in policy and practice	7
1.3 Possible concerns with current rehabilitation of grammar	11
1.4 Learners' processing of input: An under-researched area, particularly in UK	14
1.5 What can international research offer this area?: Input-based approaches to grammar pedagogy	20
1.6 A brief summary of Processing Instruction and its potential suitability for developing year 9 learners' French verb inflections	25
1.7 Summary of chapter 1	33
CHAPTER 2 THE THEORETICAL FRAMEWORK BEHIND THE PEDAGOGICAL PACKAGES: "PROCESSING INSTRUCTION" AND "ENRICHED INPUT"	34
2.1 Processing Instruction	34
2.2 Input Processing theory	38
2.3 An evaluation of Processing Instruction in terms of Input Processing theory	59
2.4 The comparative dimension: Processing Instruction versus Enriched Input	68
2.5 Summary of chapter 2 & research questions	74

CHAPTER 3 METHODOLOGICAL DISCUSSION AND THE CURRENT STUDY	76
3.1 A critical review of literature regarding the role of an experimental approach in educational research	76
3.2 The current study: The participants, their settings and the build up to the experimental intervention	88
3.3 The quasi-experimental design during the intervention	93
3.4 The language sample for the teaching materials and tests	98
3.5 Designing the tests	100
3.6 The language tests	105
3.7 Monitoring the intervention process	110
3.8 Procedures for the analysis of the achievement tests	113
3.9 Summary & original features of the study	123
CHAPTER 4 THE LEARNING CONTEXTS AND PARTICIPANTS' ATTITUDES	126
4.1 Portraits of the lessons	127
4.2 Assessing the similarity of the classes at the outset	144
4.3 Analysing the 'attitude data'	147
4.4 Summary of chapter 4	160
CHAPTER 5 RESULTS AND ANALYSIS OF THE INTERPRETATION LANGUAGE TESTS	162
5.1 Analysis of the listening tests	162
5.2 Analysis of the reading tests	179
5.3 Summary of chapter 5 and discussion	188
CHAPTER 6 RESULTS AND ANALYSIS OF THE PRODUCTION LANGUAGE TESTS	191
6.1 Analysis of the writing tests	192
6.2 Analysis of the speaking tests	203

6.3 Summary of chapter 6 and discussion	216
CHAPTER 7 SUMMARY AND CONCLUSIONS	219
7.1 Summary of the study	219
7.2 Original features of the study	220
7.3 Findings	221
7.4 Discussion of the findings	222
7.5 Limitations and further implications for future research	227
 List of references	 230
 List of appendices	 253
 Appendices	

List of figures

Figure 1.1 Example of structures box, McNab 1994 (p44).	4
Figure 1.2 Extract of observation notes taken from an action research project focussing on teaching French present and perfect tenses in year 9	18
Figure 1.3 Extract of observation notes taken from an action research project focussing on teaching and learning of German word order in year 9	19
Figure 1.4 Model of language learning often used in Input Processing literature.	27
Figure 2.1 A comparison of PI and the output / traditional instruction in PI studies	69
Figure 2.2 Comparison of PI affective activities with equivalent EI activities	74
Figure 3.1 The Experimental Design.	94
Figure 3.2 The timescale of the experimental intervention	103

List of tables

Table 3.1 Numbers from each sample taking oral test tasks	109
Table 4.1 One way ANOVAs to compare the pre test scores in all measures in different classes	145
Table 4.2 Post hoc tests following one way ANOVAs to compare pre test scores in all measures in different classes	145
Table 4.3 Results of the Likert scales regarding pupils' reported preference for learning French	149
Table 4.4 Friedman tests to find differences in learners' reported preference for learning French	150
Table 4.5 Two-tailed Wilcoxon tests to compare differences between pairs of tests in pupils' reported preference for learning French	150
Table 4.6 Results of the Likert scales for learners' reported opinions regarding the difficulty of French	152

Table 4.7	Friedman tests to find differences in learners' reported perception of difficulty of French.	153
Table 4.8	Summary of results from pupils' questionnaires regarding the PI and EI intervention materials	155
Table 5.1	Effect sizes of gains between listening tests	166
Table 5.2	Repeated measures ANOVA of the listening test results	168
Table 5.3	Planned contrasts of listening tests	168
Table 5.4	Three separate repeated measures ANOVAs on separate classes for listening tests	170
Table 5.5	Planned contrasts in each separate class for listening tests	170
Table 5.6	One sample t tests to explore the possibility that learners were using random selection in the listening test.	174
Table 5.7	Independent samples t test to compare the gains made by classes A and C between the pre and post listening tests	177
Table 5.8	Effect sizes of gains between reading tests	181
Table 5.9	Repeated measures ANOVA for the reading scores	183
Table 5.10	Planned contrasts of reading tests	183
Table 5.11	Extract from repeated measures ANOVA showing the effect of group on the transformed reading scores	184
Table 5.12	Extract from the planned contrasts on the transformed reading scores	184
Table 5.13	Three repeated measures ANOVAs on the separate classes for the reading tests	185
Table 5.14	Planned contrasts in each separate class for the reading tests	185
Table 5.15	Independent samples t test to compare the gains made by classes A and C between pre and post reading tests	187
Table 6.1	Effect sizes for gains in writing scores	196
Table 6.2	Repeated measures ANOVA for the writing scores	196
Table 6.3	Simple and repeated contrasts of the test variable for the writing scores	197
Table 6.4	Repeated measures ANOVA on the square root of the writing scores	198
Table 6.5	Simple and repeated contrasts of the test variable for the square root	

of the writing score	199
Table 6.6 Repeated measures ANOVA for class A, B & C's writing test scores	200
Table 6.7 Simple and repeated contrasts for class A, B and C's writing test scores	200
Table 6.8 Independent samples t test to compare gains made by classes A and C between pre and post writing tests	202
Table 6.9 Total scores available for oral production tasks	204
Table 6.10 Correlations between the sentence and narrative oral scores	206
Table 6.11 Paired samples t tests to compare the sentence and narrative oral scores	206
Table 6.12 Independent sample t tests to compare the oral sentence level scores of learners who took only the sentence level tasks and learners who took both the sentence and narrative level tasks	207
Table 6.13 Effect sizes of gains between speaking tests	209
Table 6.14 Repeated measures ANOVA for speaking test scores	210
Table 6.15 Simple and repeated planned contrasts between speaking tests	211
Table 6.16 Repeated measures ANOVAs carried out on classes separately for the speaking tests	213
Table 6.17 Simple and repeated planned contrasts after repeated measures ANOVAs on separate classes for the speaking scores	213
Table 6.18 Independent samples t test to compare the gains made by classes A and C between pre and post speaking tests	215

Acknowledgements

I am very grateful for the time and effort spent by the teachers and learners who participated in this study. Their enthusiasm and patience were much appreciated.

I would like to thank Professor Ros Mitchell for her exceptional supervision throughout the work for this thesis and for her role in my general education over the last few years.

I would also like to thank my parents, for their continued interest and encouragement. Particular thanks to Mum and Anne for keeping Katie so happy while this thesis was being finished.

Greg has, as always, shown great personal support and an apparently unbendable ear, but I also thank him for 950 ‘vous, il, nous, ilS’, for proof reading this entire thesis and for NOT refusing to talk about statistics.

And this is for Katie – for being the most welcome interruption imaginable.

Introduction: Overview of the thesis

This thesis presents a quasi experiment which investigated the effects of Processing Instruction (a type of input-based grammar pedagogy) on learners' ability to interpret and produce a range of French verb inflections. These effects were explored by comparing them to the progress made by two other sets of pupils: pupils experiencing another input-based approach (Enriched Input, which contrasted with some key features of Processing Instruction), and pupils in an intact class with no experimental intervention. This introduction briefly describes the layout of the thesis.

In the first chapter, the rationale is presented, justifying the need for innovation in grammar teaching in UK key stage 3 Modern Foreign Language (MFL) teaching, in particular to promote the ability of learners to manipulate verbs. The choice of two input-based approaches (Processing Instruction, PI and Enriched Input, EI) is also justified, by suggesting their potential contribution to current MFL teaching. Chapter 2 presents PI in detail as well as the theoretical framework underlying PI, namely Input Processing (IP) (VanPatten 1996, 2002 and 2004). PI is also evaluated in terms of IP theory. The theoretical and pedagogical rationales behind the comparative intervention, EI, are laid out. The research questions are provided at the end of chapter 2. The third chapter opens with a discussion of the use of experiments in educational research, which influenced the design of the current study. The methods adopted for carrying out the quasi experimental design, a description of the study and the statistical procedures used for analysing the data are then presented. Chapter 4 presents several analyses assessing and enhancing the validity and reliability of the experimental design: the three different classroom contexts are described (mainly using data from observation records); the comparability of the three classes' scores at pre test is assessed; and the data collected regarding the pupils' and teachers' attitudes to the experimental intervention is discussed. The results from the pre, post and delayed post language tests are analysed in chapter 5 (the listening and reading tests) and chapter 6 (the writing and speaking tests). In conclusion, chapter 7 summarises the study and reviews the explanations for the findings, particularly for the differing results from each school and the implications these have for the PI, IP and educational research agendas.

Chapter 1 Rationale

Introduction

This chapter sets out the educational reasons for undertaking an investigation of innovative approaches to grammar pedagogy, and in particular input-based approaches, in order to enable learners to manipulate inflectional verb morphology.

Section 1.1 provides a summary of the problems and dissatisfaction that have been experienced regarding UK secondary school foreign language learners' ability to manipulate language creatively (and in particular, verbs). Section 1.2 presents a summary of key policy changes and grassroots initiatives to reinstate grammar into the Modern Foreign Languages (MFLs) curriculum. Section 1.3 suggests some of the shortcomings of these innovations and leads to, in section 1.4, the suggestion of one area that has not yet received significant attention, namely what learners are asked to do with the target language input¹ they receive. This is illustrated with some examples of current practice. Section 1.5 discusses international research findings related to input-based approaches to grammar pedagogy. Section 1.6 presents a particularly promising pedagogical and theoretical framework for the current study (PI) and discusses the main reasons why PI seems to be worthy of attention specifically in terms of its potential role in Key Stage 3² MFL teaching. The chapter is summarised in section 1.7.

1.1 Dissatisfaction with the teaching and learning of 'language form'

Grammar pedagogy³ in UK Modern Foreign Languages (MFLs) teaching has had variable status over the last few decades, ranging from its prominence in the grammar translation methodologies, prevalent when MFLs were taught only at grammar schools, to a very

¹ The term 'input' refers to any target language (oral and written) with which learners come into contact (see discussion in section 2.4.3 regarding the difficulty of distinguishing 'input' from 'output').

² Key Stage 3 includes years 7 to 9 of the education system in England i.e. 11 – 13/14 year olds. This is the only time that learning an MFL is obligatory in England.

³ In this chapter, the terms explicit grammar teaching, grammar pedagogy, grammar teaching, the teaching of forms & structures are all used interchangeably and refer generically to aspects of language instruction where teachers intentionally draw learners' attention to the morphosyntactic structure of language.

reduced status in the late 1970's, 80's and early 90's. A full account of this history is not relevant to this thesis and can be found in Hawkins (1996), Boyd (2001), Mitchell (1988, 2000a) and Wringe (1996). Suffice to note that grammar pedagogy has been a controversial area in MFLs teaching in the UK (and elsewhere) for many years. Current methodologies share many features of functional / notional / situational / transactional language teaching approaches developed during the 80's and 90's. This section documents a range of potential concerns associated with recent and current grammar pedagogy and early learners' ability to manipulate language creatively. A wide range of sources is used, including government inspectors' reports, local level action research, pedagogical guides, text books, academic research from applied linguists and teacher trainers. They are unified by the fact that they all provide a motivation for investigating innovative approaches to teaching language form to early language learners.

1.1.1 Routinised formulas

One of the major characteristics of recent and current practice in UK schools is extensive presentation and practice of pre-assembled phrases with one-to-one equivalents of English phrases (e.g. Mitchell 1988). External assessments for beginners (Standard Attainment Tests (SATs) and General Certificate of Secondary Education examinations (GCSEs)⁴) require learners to produce accurate and fairly complex language that is also fluent. Rote-learned phrases, although a heavy burden on the memory, are an efficient means of achieving success at this level. The pervasiveness and considerable durability of these 'chunks' in learners' language has been well-documented in process studies (Adinolfi 1994, Mitchell & Martin 1997, Page 1999, Howard 1999). It has been suggested (Myles, Hooper & Mitchell 1998, Myles, Mitchell & Hooper 1999), that they may interface with and aid (and possibly be essential for) the learners' own creative use of language.

However, such extensive phrase-book learning is widely-considered as superficial, both in terms of content and lack of autonomy it gives the learner to create new language (e.g. Mitchell 1996, Chambers & Richards 1995). The process of analysing the formulas and the gradual spread of use of their component parts has been shown to be slow and the chunks

can remain unanalysed for many years (Myles 2004, Page 1999, Howard 2002). One challenge that faces teachers, therefore, is how to help learners to capitalise on and yet also move away from these routinised formulas. For example, a recent teacher debate on linguanet forum⁵ was concerned with preventing learners from using *je* and *j'ai* interchangeably⁶.

1.1.2 Bias towards nouns

UK teaching process studies (e.g. Mitchell & Martin 1997) and some popular text books⁷ demonstrate the preponderance of the teaching and learning of nouns, and to a lesser extent, adjectives. Very often one fixed subject + verb combination e.g. *avez-vous, je voudrais, j'ai, il est, il y a* is presented with a wide selection of possible complement slot fillers to suit particular situations e.g. pets, groceries, personal descriptions. A very salient characteristic of most text books is the heavy use of 'selection boxes' (see *Figure 1.1*) which give learners prefabricated verb phrases with choices of alternative adjectival / noun phrases, though sometimes a range of inflected verbs is offered.

Eric	habite	dans le nord de la France en Afrique	à Paris aux Antilles	au Lavandou au Canada
En hiver En été	il pleut	beaucoup	souvent	
	il fait	assez / très froid	assez / très chaud	
	il y a	du vent	du soleil	

Figure 1.1 Example of structures box, McNab 1994 (p44).

Using the box to write encourages learners to comprehend the choice of complements on offer but does not *require* them to understand the verbs (except 'il pleut') (nor carry out any morphosyntactic analysis, though this is discussed in section 1.4). A bank of nouns / adjectives with some adverbs but few verbs does not empower learners to create language

⁴ SATs are for year 9 learners (generally after 3 years of MFL learning) and GCSEs are for year 11 learners (generally after 5 years of MFL learning). They include listening, reading, writing and oral tests.

⁵ <http://www.linguanet.org.uk/forum.htm>

⁶ Solutions were offered along the lines of heightened metalinguistic awareness and output practice, but these issues are discussed in more detail later.

⁷ see the glossaries in Avantage, Etoiles, Tricolor, Logo, Route Nationale, Télescope, Arriba, Auf Deutsch.

for their own purposes (Lee 1998, McPake 1999, Chambers 1999). The central role of verbs in driving language acquisition is widely accepted, both in first (native) and second language (L1 and L2) acquisition (Aitchison 1998 (p161), Harley 1986, Little & Singleton 1991 (p131)). Verbs appear to be at the root of some of the most salient and persistent non-target-like phenomena in interlanguage⁸ grammars in naturalistic, immersion and classroom learners, regardless of one's perspective (e.g. Hawkins 2002 (a Universal Grammar (UG) perspective), Harley & Swain 1978 (a cognitive perspective with immersion learners), Dietrich, Klein & Noyau 1995 (a functionalist perspective with naturalistic learners), Howard 2002 (an aspect hypothesis perspective amongst UK university learners) and Dobson 1998 (a pedagogic descriptive approach with UK Key Stages 3 and 4 learners)).

1.1.3 Bias towards first person

A number of studies (e.g. Myles, Mitchell & Hooper 1999, Grenfell 2000, Page 1999) attest to the preponderance of routinised dialogues, featuring 'je' (particularly *j'ai*, *j'aime* and *'je m'appelle'*) and 'tu' (particularly *'tu as'*, *'tu aimes'* *'tu t'appelles'*). It could be argued that there is, in any case, a bias in natural languages towards first and second person reference which gets incorporated into grammars: "in oral discourse the topic is also particularly likely to be the speaker or the listener. This social universal results in a statistical bias towards first and second person subjects" (Bates & MacWhinney 1989, p21). However, the teacher and learner process studies, cited above, indicate that the bias is not only particularly heavy in classrooms but it also focuses on *singular* first and second person. This probably contributes to the apparent difficulty learners have in referring accurately to plural subjects and third person subjects.

1.1.4 Dissatisfaction amongst stake-holders

Poor linguistic accuracy during, and following, the 11-16 phase is a recurring complaint in the literature (Wright 1999, Wimpory 1997, Hurman 1992, Metcalfe, Laurillard & Mason 1995, Klapper 1997, Gray 1999, Neil & Laverty 2001, Dobson 1998, <http://www.lang.ltsn.ac.uk/lings/linggrammarlinks.html>). The annual CILT (now the

⁸ The term 'interlanguage' (IL) is used throughout this thesis to refer to a learner's language system which is not yet native-like (see also section 3.6.5 and footnote 33).

National Centre for Languages) Languages Show provides workshops such as 'Bridging the GCSE-AS-A level Gap' and 'Dealing with Grammar'.

OFSTED⁹ subject reports (OFSTED 1995, 2001& 2002 and Dobson 1998) give us some indication of the government inspectorate's dissatisfaction with MFLs learners' ability to manipulate form:

"Many pupils need to develop ... their confidence in speaking and the accuracy of their writing... more improvement is needed, for example in ...the accurate application of grammar." (OFSTED 2002)

One of the most recurring problems is perceived as being learners' lack of ability to cope with unfamiliar situations or create their own language:

"Some aspects of the programme of study, such as those requiring pupils to take the initiative or cope with the unfamiliar, remain underdeveloped" (OFSTED 2001)

Dissatisfaction with progression in Key Stage 3 in particular has been reported:

"After a good start in Y7, many pupils lose momentum... As a consequence, many pupils have an insufficient basis for progression through Key Stage 4 and beyond. This is particularly the case where their understanding of how language works, grasp of basic grammar and powers of recall are weak. Examples include a narrow range of adjectives and limited ability to manipulate the tenses of common verbs. In such cases pupils struggle to cope with more demanding tasks, particularly where these involve taking the initiative or coping with the unpredictable... the progress of pupils at Key Stage 4 has improved but is held back where sufficient momentum for learning has not been generated in Year 8 and Year 9" (OFSTED 2001).

Pupil dissatisfaction

There is also evidence to suggest that pupils are aware of their lack of grammatical autonomy (Lee 1998, McPake 1999, Chambers 1999, Harris et al. 2001). Stern, Burstall & Harley (1975) gave some indication of the cyclical nature of achievement and motivation, suggesting that when learners did not feel satisfied with their progress their motivation waned and this affected further achievement. An additional impetus to investigate alternative types of grammar instruction is that extrinsic motivation for learning MFLs is weak in the UK (partly because of global English). In a different sociolinguistic context,

⁹ Office for Standards in Education: the Government's inspectorate of schools.

with more instrumental motivation and / or fewer 'affective filters' (Krashen 1982), details of teaching techniques may have less impact on learners' attitudes or progression whereas in MFL classrooms in L1 English contexts it has been suggested that such 'details' may have more influence (Dörnyei 2001).

1.2 Current rehabilitation of grammar in policy and practice

This section provides a general picture of how grammar pedagogy is broadly conceived in recent government and teacher-led initiatives, which partly developed as a result of the concerns laid out above. The section also presents a brief description of common verb related grammar activities in popular text books.

1.2.1 Government initiatives involving the conceptualisation of grammar pedagogy

The dissatisfaction with pupils' ability to create language, outlined above, has been accompanied by (and has partly led to) a return to traditional views of literacy and grammatical accuracy promoted via central government policies. These have revived an interest in more explicit grammar instruction in MFL and English teaching in UK schools. The DfEE response to the Nuffield Enquiry (DfEE 2001) suggested that Specialist Language Colleges, the National Curriculum for Modern Foreign Languages in England and Wales (DfEE 1999), the Scheme of Work (QCA 2000b), the GCSE Criteria (QCA 2000a) and the Framework for Teaching MFLs (DfES 2003) provide some indication about what teachers, according to most recent policy, are to consider good practice in terms of grammar pedagogy, "with more emphasis on knowledge, understanding and skills" (DfEE 2001). These developments are briefly described in appendix 2.

In summary, the National Curriculum (NC) provides a general indication to teachers that language form should be taught and that learners benefit from explicit knowledge of form. The NC levels of achievement and GCSE criteria and examinations show that there is a continued (and slightly increased) emphasis on productive accuracy, involving assessment via traditional verb paradigm type activities. The Key Stage 3 Scheme of Work offers more precise advice regarding specific activities designed to promote learners' use and

awareness of particular language forms, though these appear to be a largely untapped resource¹⁰. The Key Stage 3 Framework, non-statutory guidance which has just become available for use, has a heavy emphasis on automatization of explicit knowledge about language.

1.2.2 Teacher-led initiatives involving grammar pedagogy.

In practice, grammar teaching in schools probably remains largely context-(i.e. teacher-, school-, text book-) dependent (Mitchell 1994, Bygate, Tonkyn, & Williams 1994b)¹¹. Since the policy documents above have been published, I could not locate any refereed studies documenting current grammar pedagogy. Nevertheless, it has been suggested that in order to focus learners' attention on form, teachers tend to use activities with roots in grammar-translation methods (e.g. the teaching of whole paradigms, "explanation, exemplification, and drawing parallels with English/translating" (Mitchell & Brumfit 2001, p285) or in behaviourist audiolingual learning techniques (e.g. drill or 'slot & fill' type activities). These activities are often packaged more attractively than in previous methods by using colour coding, games, actions, mimes and peer correction as seen in Wells (2000) and some of the recommendations in pedagogical guides (e.g. Harris et al. 2001)¹².

Examples of recent teacher initiatives in grammar pedagogy illustrating some of these characteristics can be found in six government-funded action research projects in 2000/1 addressing the broad field of grammar pedagogy¹³. They give a broad indication of a grass-roots movement which reflects a renewed interest in grammar instruction, suggesting there is a feeling that 'grammar is back' (e.g. <http://www.linguanet.org.uk/forum.htm>, <http://www.languagelearn.co.uk/>). Four extracts from observation records from some of these projects, in which I was involved as a research assistant, are discussed here¹⁴.

¹⁰ Toye (2001) & Johnston (2001) (action research projects, see http://www.teachernet.gov.uk/Professional_Development/opportunities/bprs) and casual enquiries suggest that they are used very little.

¹¹ The status, quantity and techniques of grammar teaching are usually determined by many experiential, psychological, professional and contextual factors e.g. individual teachers, departments, textbooks (Mitchell & Hooper 1992, Mitchell & Brumfit 2001, Mitchell 2000, Borg 1999).

¹² This also discusses how to incorporate discussion of form into more spontaneous or meaning-oriented activities.

¹³ http://www.teachernet.gov.uk/Professional_Development/opportunities/bprs.

¹⁴ To maintain anonymity, these observation records are not referenced. See appendix 1 for the extracts discussed.

Although interpretation of this small selection is cautious, the projects were carried out in a Language College which is respected locally for its MFL and overall teaching.

One action research project focussed on the teaching and learning of present and perfect tense verb inflections in French. One lesson (extract 1 in appendix 1) began with a whole class oral production activity where learners had to follow a predictable pattern (repeating the phrase 'je suis allé au cinema' adding a time adverbial). The teacher provided model answers and used translation and transliteration to explain the task. She insisted on accurate production, using explicit error correction. The learners' responses indicated that they were aware that corrections referred to form (not content). Extract 2 (appendix 1) illustrates a whole class oral activity (in the style of the game show "Blockbusters") which required learners to translate into French verbs in the perfect tense with a range of subjects and provide the orthographic form where this was not evident from the spoken form. At the end of the lesson each learner was required to say either 'nous regardons' or 'nous avons regardé' depending on whether the teacher said 'present' or 'perfect'.

Another action research project focussed on the teaching and learning of explicit knowledge and ability to use present tense verbs in German. Extract 3 in appendix 1 describes a whole class oral activity, practising a pattern of replacing 'ich + verb' with 'sie / er + verb ending in t'. There was an emphasis on accuracy and extensive error correction. There were few comprehension checks regarding the meaning of the forms being practised. The teacher points out explicitly that learners often confuse 'gern' and 'gehen' and provides the correct syntax. Extract 4 in appendix 1 demonstrates a fairly abstract / analytic / deductive approach to grammar pedagogy, with the use of full verb paradigms, translation, metalanguage, explicit description of rule systems (L1 and L2) and decontextualised examples with little reference to meaning. The observer commented that pupils may have relied on strategies other than comprehension of the verbs to complete the tasks and that several pupils still had trouble identifying verbs by the end of the lesson.

These examples go some way to corroborating research, carried out prior to the recent policy changes, which found that grammar pedagogy often manifests itself in techniques

reminiscent of grammar translation and audio-lingual styles (e.g. Johnson 1994, Brumfit, Mitchell & Hooper 1996 and several studies in Bygate, Tonkyn & Williams 1994a)¹⁵.

1.2.3 Published material

There is considerable reliance on textbooks and published materials in school MFLs lessons (Lee 1998, Mitchell & Martin 1997, OFSTED 2001, Mitchell & Brumfit 2001). Therefore, a brief look at what they offer in terms of the teaching and learning of verbs is in order¹⁶.

A common feature is that text books provide short summaries of grammar rules usually at the back but sometimes within the body of the book. This would indicate that most learners do have some exposure to rules of language, although the extent to which their attention is directed to them depends on the teacher and the learner. Mnemonics are often provided to aid in the decision as to which auxiliary to use in the perfect tense ('être' or 'avoir').

Most verb-related grammar exercises (including website and Computer Assisted Learning materials) are usually written, though often oral, and frequently involve entire verb paradigms. They are sometimes associated with an NC or GCSE topic area e.g. describing your family. Activities include: matching an inflected verb with its infinitive or vice versa, filling in a blank with a given inflected verb (both of which can usually be done by using the verb stem alone) or providing the verb inflections according to the given subject, which is usually a pronoun (and according to the given verb, in the perfect tense)¹⁷. Learners are also asked to note instances of a particular grammatical feature (e.g. by underlining) or to provide an account (usually written) referring to their own lives (e.g. their hobbies, holidays) following a model (e.g. McNab 1994, p69).

¹⁵ It has been documented that these techniques are/have been particularly used for teaching the present and perfect tenses (e.g. Macrory & Stone 2000, Page 1999, Graham 1997 and also demonstrated in many published materials), and were probably popular even during the most transactionally-oriented periods of language teaching.

¹⁶ A selection of common text books is used here (Avantage 1 & 2, Etoiles 1 & 2, Tricolor 1, 2 & 3, Logo, Route Nationale 1, 2 & 3, Télescope 2, Arriba 1, 2 & 3, Auf Deutsche 1 & 2).

¹⁷ See for example <http://www.quia.com/jfc/1095.html>, <http://www.didieraccord.com/exercices/1.2.3.html>

Another feature shared by most textbooks is that they move swiftly between different forms. For example, in Taylor & Edwards (1994) activities on p33 are devoted to the expression 'j' ai peur + du / de la / des', on p34 to presenting and practising 'first person verb + infinitive' and p35 to 'regular and irregular forms of the perfect tense'.

This brief review of recent studies documenting teaching and currently teaching materials has suggested that a wide range of grammar teaching techniques are used, frequently consisting of a mix of metalinguistic grammar translation activities (i.e. encouraging learners to automatise explicit rules) or behaviourist slot filling production activities in writing and speaking.

1.3 Possible concerns with current rehabilitation of grammar

This section sets out the principal concerns with the developments described above. These concerns point to the need for research into innovative and effective approaches into how to incorporate grammar pedagogy into broadly communicative teaching.

1.3.1 *Lack of engagement with international Second Language Acquisition and Focus on Form/S research*

Many applied linguists indicate the importance of improving the interface between practice and international studies regarding MFL pedagogy e.g. Hudson (2001), Grenfell (2000), Mitchell (2000), Allford (2003). Mitchell (2003), Boyd (2001), Pachler (2003) and Marsden (2001) suggest that many aspects of the National Curriculum (NC), GCSE specifications and the new Key Stage 3 Strategy could be better supported by SLA research e.g. the expected routes of progression, and guidance about whether or how to incorporate the teaching of form into a communicative curriculum¹⁸.

¹⁸ The NC programme of Study (PoS) only states "Pupils should be taught the grammar of the target language and how to apply it" (PoS, 1,b, DfEE 1999). 3b also suggests, as a non-statutory example, "pupils should be taught how to use the context and other clues to interpret meaning [for example, by identifying the grammatical function of unfamiliar words...].

There is a body of SLA literature which offers a range of theories and pedagogical techniques regarding how and when to teach grammar (Long & Robinson 1998, Norris & Ortega 2000, Spada 1997). Focus on Form and FormS (henceforth FonF/S¹⁹) research has not received a great deal of attention in the UK secondary school context (Mitchell 2000)²⁰. It comprises a largely neglected body of research which may offer some framework for systematic investigations in grammar pedagogy in UK classrooms.

1.3.2 Disadvantages of approaches purely based on grammar translation methods

As described above, grammar pedagogy techniques such as explicit grammar rules, mechanical practice reminiscent of behaviourist approaches (e.g. slot replacing) and grammar translation activities (including metalinguistic and explicit error correction strategies) are common ways in which the revival of grammar pedagogy in UK schools appears to manifest itself²¹. This section suggests some of the reasons why a re-adoption of grammar-translation and/or behaviourist techniques alone may not be satisfactory.

a) Grammar 'versus' communication

Grammar translation or behaviourist approaches (even current, more appealing forms of these) can often mean that grammar pedagogy is held largely 'in opposition' to more communicative or functional activities. The need for pedagogical guides such as Harris et al. (2001) suggests that incorporating grammar into classroom activities without resorting to thinly disguised grammar translation activities is difficult. Mitchell & Hogg (2001) found that there was a tension between certain FonF/S techniques and other classroom activities (e.g. practising vocabulary, transactional phrases, target language use).

¹⁹ Focus on Form tends to refer to more implicit / communicatively embedded / task essential / reactive grammar pedagogy techniques. Focus on FormS tends to refer to techniques which are synthetic / proactive / involve abstract practice. Therefore, the term FonF/S is used throughout this thesis to encompass all such techniques (as in Norris & Ortega 2000), to refer to "any pedagogical effort which is used to draw the learners' attention to language form either implicitly or explicitly" Spada (1997) (p73) (though Spada used the term "Form Focussed Instruction").

²⁰ though see Grenfell & Harris (1999), Macaro (2001a & b) Coyle (2001) and Klapper (2003) in the areas of learning strategies, role of L1, interactionist perspectives and task-based learning respectively.

²¹ There are few examples of a push for a focus on meaning in school MFL teaching in the UK, reviewed in Mitchell (2002) and Grenfell (2000) (e.g. Clark 1988, Charis 1996, Coyle 2000 & 2001, Harris et al. 2001). Although a promising programme of innovation and research, they remain fairly small-scale when compared, for example, to the piloting and implementation of the Key Stage 3 strategy, the impact of text books and washback from GCSEs and the NC.

Lightbown (2001) reports a more exaggerated manifestation of this in immersion programmes²² - learners are either taught "important academic content" or, in language classes, offered "a time to do conjugations or grammar exercises" p91.

b) Appeals to an elite

Experience of grammar teaching is very often limited to 'top ability' classes (Wells 2000, Metcalfe 1997, Rendall 1998, Mitchell & Brumfit 2001). Some of the current techniques (i.e. heavy reliance on written forms, complex rules, rote-learning and long term memory capacity) may be more suited to particular learning styles (i.e. academically oriented learners), as acknowledged by DeKeyser 1998 (p62), one of the proponents of proceduralising declarative knowledge in language learning. It has not been shown that academically oriented techniques are the most appropriate tool in the context of MFLs for all at Key Stage 3.

c) Compromises creativity

Hawkins (1996) suggests that experience of grammar translation methods in UK schools resulted in poor oral fluency and low motivation. Lightbown & Spada (1993) suggest that an extreme form of 'getting it right from the beginning' does not benefit effective communication (p79-83).

d) Weak link between language knowledge and use

It has been shown that the relationship between explicit knowledge of rules and actual language use is far from clear cut. Alderson, Clapham, & Steel (1998), Green & Hecht (1992), Page (1999), Brumfit, Mitchell & Hooper (1996) and Metcalfe (1997) have all shown that learners' knowledge, or at least their ability to articulate it, remains hazy. The first three studies also suggested that learners' ability to describe rules of language accurately does not reflect their ability to use those rules (also see Terrell, Baycroft, & Perrone 1987, Scott 1989). VanPatten & Oikkenon (1996), Sanz & Morgan-Short (2004) and Benati (2004) have suggested that explicit knowledge alone does not lead to significant learning gains. Truscott (1998) and Schwartz (1993) review studies that have shown that the effects of explicit knowledge are short-lived.

²² Immersion programmes offer part or all of the curriculum through the medium of an L2.

e) *Tendency to ignore the learner's 'internal syllabus'*

One criticism of grammar translation and behaviourist techniques is that they do not always heed the well documented phenomena of fairly fixed developmental routes (Ellis 1994, Larsen-Freeman & Long 1991, Sanz 1999, Long & Robinson 1998). Teachers often acknowledge that forms are explicitly presented and practised which are probably beyond the cognitive or linguistic stage of the learners²³. Learners can be asked to produce forms which they are not able to process²⁴ from the input. Issues relating to the use of input are introduced in the next section, in particular the potential role of input-based FonF/S approaches to constructing grammar teaching activities.

1.4 Learners' processing of input: An under-researched area, particularly in UK

This section discusses and supports the claim that grammar pedagogy rarely takes the form of input-based activities and that current listening and reading activities do not require learners to comprehend certain aspects of language form.

1.4.1 *Input-based activities under-represented in grammar pedagogy*

What learners do with the input they tend to receive in UK MFL secondary education has been given little attention, yet it is often argued that processing input is probably the most vital component of SLA (Chaudron 1985, Gass 1988, Krashen 1982, Carroll 2001). Lightbown (1992) and many studies of input-based approaches (discussed below) suggest that certain types of input practice contribute towards improving the accuracy of the developing system.

Given the emphasis most models of language learning put on comprehensible input, listening and reading are perhaps under-represented skill areas in UK classrooms. The KS3 schemes of work (QCA 2000b) suggest a great deal more production than comprehension activities. The five classes discussed in Mitchell & Martin (1997) had listening activities

²³ as in some post observation teacher interviews I carried out in the action research projects discussed in section 3.2.

on average 7.7% of class time and reading 4.1%, giving an average total of 'pure input' time of 11.8%. Although this discounts the input provided by whole class speaking (62.4%) and paired speaking activities (7.6%), these activities were commonly found to rehearse fairly routinised patterns as discussed in sections 1.1 and 1.2.2, and do not force learners to analyse input in any great detail (e.g. verb inflections).

Despite attempts to increase the use of target language in the classroom, research suggests that learners are usually not *required* to understand much teacher or pupil target language (TL) (Dickson 1996, Mitchell 1988). This is often because either the TL is comprehensible using other means (such as the context, paralinguistic communication, pupil or teacher translation) or because TL comprehension is not necessary to function as a pupil in that class (target language is not often used for essential classroom administrative / management).

Of most relevance to this study is that grammar instruction is predominantly perceived as production practice, with some use of reading tasks (Ellis 1999, VanPatten 1996, Paulston 1972). Though these works do not refer specifically to the current UK context, many other sources suggest there is limited evidence of 'reading grammar' and less, or none, of 'aural grammar' (i.e. micro skills such as interpreting the meaning of grammatical form). This is supported by process studies (Mitchell & Martin 1997, Mitchell & Hogg 2001), by grammar pedagogy guides and research (Rendall 1998, Wells 2000) and by pedagogical guides regarding listening activities (Turner 1995).

1.4.2 Input-based activities in practice

This section argues that listening and reading activities rarely force learners to interpret the meaning of all aspects of the input. Two brief presentations of evidence for this are given: first a few examples of a popular textbook's listening and reading activities and secondly some excerpts from classroom observations, including teacher-designed listening and reading activities and target language classroom interaction. The sample of activities and observations presented here is small and no particular theoretical framework is adopted.

²⁴ 'process' refers to attending to and detecting (i.e. interpreting) input. These issues are discussed in more detail in chapter 2.

The aim is to explain in broad, non-technical terms one of the principal rationales of the study: namely a concern that the way learners are required to process input does not promote attention to specific features in the language and in some cases, this is contrary to the intention of the teacher or text book. A more theoretical approach to the issues is presented in chapter 2.

a) Examples of listening and reading activities in published teaching materials

- i) In McNab & Barrabé (1993) p106, learners are asked to write whether they heard 'du', 'de la', 'un' or 'une' in an audio-recording of phrases containing these forms, though they do not have to decide which goes where, understand why they are different or what nouns must follow. Perhaps such activities raise awareness that different forms exist, but it seems likely that learners would attend to the surface phonological forms only, without further interpretation of different meanings / uses of the forms in the input.
- ii) In McNab 1(994) p22 (see appendix 3a), the focus of the exercise is 'tutoyer ou vouvoyer' and learners must decide which the speaker on the tape does by marking T or V. In order to do this, the learner can complete the activity after noticing only whether they heard 'tu' or 'vous', and any attention given to the different verb inflections would be incidental to the task. The next activity requires learners to replace, in writing, 'tu + verb forms' with 'vous + verb forms' and vice versa, though verb forms required are provided for them. After another activity containing examples of present tense 3rd person plural verbs, learners have to produce a range of questions to their friends using 'tu'.
- iii) In McNab (1994) p23 (see appendix 3b), the final production activity in this unit requires learners to write 3rd person present tense verbs in a report of a class survey (e.g. '10 personnes aiment le football'). However, the listening and reading activities offered in preparation for this do not require learners to attend to the verb inflection²⁵. The listening activity requires learners to understand only numbers (e.g. ' ___ % travaillent plein temps'). The reading activity requires learners to copy inflected verbs in their entirety without

²⁵ Clearly the oral inflection is null, but the point remains that learners do not have to attend to this by, say, contrasting it with a distinguishable inflection.

separating the lexical verb and inflection (the rubric, 'Nombre de mères qui...', tells the pupils they are referring to 3rd person plural).

- iv) None of the verb paradigm activities mentioned in section 1.2.2 and 1.2.3 require learners to 'interpret' the verb inflection for its meaning in terms of tense, person or number in order to assign the suitable subject (instead they ask learners to interpret the subject in order to select the inflection).
- v) In the reading activity in McNab 1994 p71 (see appendix 3c), learners can rely solely on lexical cues to decide whether the statements are true or false, partly because the text and the questions are entirely in the past and refer to one point in time. Nevertheless, the subsequent production activities require learners to write sentences in the perfect tense.

b) *Observed examples of listening and reading activities or 'events' in classrooms*

- i) In 'slot & replace' activities such as the one discussed in section 1.3, where 'ich spiele' is changed to 'er spielt', although learners must 'segment' the input (e.g. separate 'je' from 'joue' to replace 'je' with 'il'), they can usually complete such activities by following a predictable pattern, without attaching any meaning to the forms. Further examples of observed activities involving listening and reading are given here.

ii)

Written in centre of board: Appel: Quand es-tu allé(e) au cinéma?

10.15- 10.20 Pupils arriving. T asking Ps re yesterday's test; Ps asking re marks, all in English

Observer asks pupils in front what 'appel' (on the board) means - pupils suggest 'repeat' or 'answer'. They suggest the question on the board means 'what did you watch?'.

T - asks what question on board means. Les mots clés sont [underlines 3 words on board and says them:] quand allé cinéma

P1 - when do you go to the cinema?

T - repeats with rising intonation and emphasis when *do* you go to the cinema?

P1 - when did you go to the cinema?

...[T elicits a model response and elicits and provides various time adverbials then writes on board]: je suis allé au cinéma la semaine dernière, hier, le mois dernier, il y a 2 mois, 3 mois

10:20- 10:27 Ps read a response from the board when their name is called out.

Figure 1.2 Extract of observation notes taken from an action research project focussing on teaching French present and perfect tenses in year 9

In this extract the teacher directed pupils' attention to lexical items in the input ('quand', 'allé', 'cinéma'), calling these 'mots clés'. The pupil's response suggests that using this strategy alone can lead to erroneous interpretation, indicating the importance of attending to forms that are perhaps non-salient. In addition, the learners were provided with a model answer which they did not *have* to understand to produce it, whilst still ensuring success in the task. The widespread error correction strategy used in this extract ("do?") *may* have forced the learner to re-analyse the language on the board, and attend to the verb inflection 'es'. However, it may be that the learners simply guessed from the teacher's prosody and the context (e.g. the fact that learning the past tense was a prominent feature in this series of lessons) that 'did go' was the correct translation²⁶.

iii)

2.45 ... T points to and reads out the aims from the board... T reads out the title of the extract which is written on the board - 'Herr X, Ist er gesund?' and adds in English, 'you may have seen him in the canteen'. T continues - Die Antworten sind unter hier. T points to the words to fill in the gaps from the text. T - copieren. 10 Minuten, schnell. T - put chewing gum in bin. P - nicht. T - doch. P - what's doch? T - you say nicht, I say doch, no, yes. A pupil has no book and T deals with this issue in German. P - haven't got a ruler? T - was ist ruler? T gives ruler to pupil saying 'lineal'. P must repeat before he can take the ruler. Text from the board:
 ich __ ziemlich gesund. Zum Frühstück __ ich einen Apfel und zwei Stücke __. Ich __ Orangensaft und Kaffee. Zum ...

2.47 T - nicht sprechen. Schreiben. P - you have to fill in blanks. Girl near me repeats nicht [?]

2.48 My text copying continues: Mittag __ ich normalerweise in der Kantine oder ein __, wenn ich keine Zeit __. Abends __ ich am liebste Pasta und Fisch. Mein Lieblingsessen [...] __ frische Sardinen. Ich __ gern Fisch, weil es mir gut __. Ich __ gern rotwein, aber nicht zu viel. Morgens __ ich Kaffee. 'Expresso' __ lecker.

2.52 Pupils generally on task.
 P - what's the first space? P - esse.. P - what's schmeck? T - taste. P - was ist der first one? what does der first one mean? T - 5 minuten nicht sprechen. P - sir what do you have with sardines? T - shh. T - sir there's 13 words but 14 gaps. T adds another word.
 Words given on the board to go in spaces: trinke, esse, esse, bin, esse, habe, sandwich, trinke, esse, Toast, esse, ist, shmeckt, sind.

²⁶ In the same situation repeating the target language form ("es-tu allé?") or asking "does it say 'vas-tu'?" may have led learners to attend to the *meaning* of the input.

Some chatting beginning. T - einen Minute. 60 seconds (in German). P - are you going to tell us the answers? T - yes. P - that's good. Most pupils have finished copying text and starting to add missing words when the teacher stops individual work for whole class feedback

Figure 1.3 Extract of observation notes taken from an action research project focussing on teaching and learning of German word order in year 9

Figure 1.3 describes a teacher-designed written gap fill activity which was part of a series of lessons focussing on German word order, in particular subject verb inversion after adverbs and in 'weil' clauses. The missing words, mainly 1st person verbs, were provided. It can be seen that the activity can be completed by understanding the verb stem of the missing words and key lexical items in the text - verb inflections and syntax can be ignored, though learners may have processed them incidentally to the task (this is discussed in more detail in chapter 2). This extract demonstrates the use of TL for some classroom instructions and interaction although the pupils do not have to understand much of it, and certainly not verb inflections, in order to carry out the tasks required.

As mentioned in example ii), there are probably moments in lessons where learners' attention is drawn to connecting a grammatical feature in the input with its meaning. However, it is not known how frequently such moments occur. As some indication for this, in a series of 8 lesson observations I carried out for three action research projects, only one such incident was noted:

The words manqué & touché were written on board (in preparation for a game of battleships)
 T: Who can tell me whether they are present or past tense?
 P: present.
 T: past - it's got an accent on.

It is likely that questions such as these promote attention to the meaning of particular features in the input (even though it is debatable whether the participles do constitute 'past' reference).

This section has suggested that the tasks associated with oral and written input usually require learners to attend to lexical items or items such as subject pronouns, allowing

learners to be successful without attending to certain features of 'form'²⁷ in the input such as inflectional verb morphology. This may encourage a 'semantic level of processing' (Swain 1995) which may be less useful in terms of learning certain language features.

1.5 What can international research offer this area?: Input-based approaches to grammar pedagogy

Useful overviews of input-based instructional techniques and approaches to SLA are provided in Lee (1999), VanPatten (2000), Lightbown (2001), Ellis (1999), Izumi (2002) and Gass (1997), though this is perhaps a less developed area than research into output practice (Bygate, Skehan, & Swain 2001, Norris & Ortega 2000). Many input-based studies are focussed on the modification of input in interactions and some are laboratory experiments investigating the effects of implicit and explicit input modes. Other types of input-based instruction include enriched / enhanced input approaches e.g. Long (1991) or Sharwood Smith (1993), the defining feature of which is that they do not require the learner to produce the target structure. It is this latter subset of studies that was reviewed in order to look for avenues for investigating listening and reading activities for this study.

1.5.1 *Purely comprehension-based approaches*

The 'non-interventionist' approach was mainly embodied in recent years by Krashen (1985) and Krashen & Terrell (1983). In this genre learners experience the target language unmodified, as a medium of genuine communication. This is motivated by a belief that all L2 learners can subconsciously analyse linguistic input²⁸, without any intervention (including artificially providing a flood of a particular type of form in the input) to raise their awareness of the forms in the language. One of the main advantages of a purely comprehension-based approach is that it shows some respect for learners' current stage of development (Long & Robinson 1998 p17), as they are left to produce language once their own mechanisms have processed them in the input.

²⁷ a discussion of the term 'form' is provided in chapter 2.

²⁸ and / or form new neural networks and / or have access to some innate knowledge of how languages are formed

However, Krashen's input-based model of language learning has received criticism partly because the level of comprehensible input suggested by the model ('*Input + 1*') is difficult to interpret in practical terms.

In addition, there is a considerable body of evidence from immersion and naturalistic contexts (reviewed in Lightbown & Spada 1993) suggesting that although a diet of purely comprehensible input does help comprehension and production skills, it is not sufficient for learners to produce certain features in target-like ways, where low frequency in the input cannot be a cause. For example Harley (1992) found that early immersion students with approximately 1000 hours of classroom exposure to French over-generalised and used without consistency the auxiliary forms *a* and *va* and used them with non target-like past participles or infinitives. Long & Robinson (1998) suggest that this is particularly the case for items that are "rare and/ or semantically lightweight, and/or perceptually non-salient, and/or cause little or no communicative distress" (p23)²⁹.

Advantages for rate and range of structures are found for learners whose attention *is* directed towards form and more accurate production, reviewed in, for example, Bardovi-Harlig (2000), Herschensohn (2003) and Doughty & Williams (1998). Norris & Ortega (2000) conducted a quantitative meta-analysis of 49 studies and concluded that FonF/S interventions result in large target-oriented gains.

1.5.2 *Consciousness raising*

Sharwood Smith (1981 & 1993) and Rutherford & Sharwood Smith (1985) suggested that learners should be deliberately directed to attend to form by raising their awareness of certain features in the input. Fotos (1993) and Fotos (1994) used consciousness-raising (CR) techniques such as working in small groups to solve grammar problems in the TL. The studies claimed that learners are more likely to notice target features in CR tasks than when not directed in any way toward the target and that CR tasks were at least as effective in drawing learner attention to form as were those in more traditional grammar lessons.

²⁹ 'Perceptual saliency' is discussed further in section 2.2.4.

However, two problems with CR techniques are that it is difficult to ascertain whether noticing has occurred and whether such noticing constitutes the type of language knowledge used in spontaneous production or comprehension.

1.5.3 Enriched Input

Ellis (1999), in his review of input-based approaches to SLA, bunches studies which are not Processing Instruction together, labelling them 'enriched input-based studies', as they all provide the learner with "numerous exemplars of a grammatical feature known to be problematical to learners" (p68). Enriched input can take various forms: typographical markings (frequently referred to as 'enhanced input'); simple 'exposure' with or without explicitly pointing out or describing the target features (e.g. an input flood where the "principle is that the more opportunities there are in the input for learners to notice a linguistic feature, the more likely they are to do so" Doughty & Williams 1998 p236); or exposure to input with follow-up questions which direct the learner to target features. Enriched input studies have produced mixed results and the following is a brief review of studies that have examined its effects relative to other types of instruction.

A number of studies have found that enriched input can be beneficial. Jourdenais et al. (1995), Leeman et al. (1995) and Shook (1994) have found a significant effect for enhanced input for noticing target forms and, in the latter two, for subsequent output. Other studies have led to more refined suggestions about the type of linguistic feature that may be susceptible to this type of FonF/S. The experimental treatment in Trahey & White (1993) consisted of stories, games and exercises containing sentences with adverbs in the three sentence positions permitted by English. The francophone learners succeeded in learning the Subject Adverb Verb (SAV) order but failed to 'unlearn' the ungrammatical Subject Verb Adverb Object (SVAO) order. Trahey (1996) showed that these results still obtained 1 year after the intervention. By comparing with White (1991) it was suggested that the input flood was as successful as explicit instruction in teaching the SAV order but that explicit instruction appeared to be necessary to teach learners the ungrammaticality of SVAO in English. The enhancement group in Alanen (1995) did not outperform the control group in accurate use of either locative suffixes or consonant gradation rules. However, the range of target features increased in enhanced learners' productions, whereas

the control group used no such features. In Doughty (1991) one of the 3 groups, the 'meaning-oriented group' had lexical and semantic rephrasing of the relative clauses, that were also highlighted and capitalized. This group demonstrated better comprehension and a gain equivalent to that of the 'rule-oriented' group in relativization knowledge and better gains in both compared to the control 'input flood' group. However, interpretation of these results is confounded by the fact that the 'meaning-oriented group' had both rephrasing and visual enhancement focus on form techniques. Findings from Williams (1995) suggested that when the target structure is complex (relative clauses and passives) enriched input may be as or more effective than explicit instruction with feedback. However, when the target feature is simple (participial adjectives) explicit instruction may be more effective. Williams & Evans (1998) showed that acquisition of an emergent form (participial adjectives) was facilitated by an input flood (though contextualised explanations were even more helpful), whereas for a non-emergent form (the passive), the input flood seemed to help learners to notice and begin to attempt to use the form but explicit instruction was of little further use. Ellis (1993) investigated the learning of soft mutation in Welsh morphology. He found that the subjects who had both 'rules and instances [a flood of examples]' outperformed the learners given only examples and the learners given only rules.

Although the body of research above suggests enriched input can have positive effects on certain types of learning, a few studies have found that it has no beneficial impact when compared with a range of alternative interventions. White (1995) and White (1998) compared three input approaches to make English 3rd person singular possessive determiners more perceptually salient to francophone learners : a 'natural input condition', an input enhancement condition (using textual enhancement such as bolding and font size) and the input enhancement plus book flood condition (which started 2 weeks prior to the other conditions and continued afterwards). All three conditions showed some benefits and there was no significant difference between the groups. White suggests that the pretest raised awareness of all the groups, that input flooding alone may not be particularly effective and that learners may need more explicit assistance in using the input to construct the determiner system. Overstreet (1998) found that text-enhancement of a flood of exemplars interfered with comprehension. Lightbown, Halter & White (2002) found that

simply exposing learners to the target form was not as effective as explicit FonF which enabled more learners to use the pronoun system and morphological markers for past and plural. Izumi (2002) found that for the learning of English relativization those who received visual input enhancement failed to show measurable gains in learning, despite having found that enhancement had a positive impact on the noticing of the target forms in the input. Jourdenais (1998) and Leow (1997) also found no significant effect for input enhancement.

In brief, three studies have yielded positive findings for the facilitative effect of enriched input (Jourdenais et al. 1995, Shook 1994 and Leeman et al. 1995); eight studies showed limited benefits or concluded that effects were specific to certain language forms and developmental stages (Doughty 1991, Trahey & White 1993, Alanen 1995, Williams 1995, Trahey 1996, Robinson 1997, Williams & Evans 1998 and Williams 1999), and seven studies found no significant effect at all, compared with a range of alternative treatments (Overstreet 1998, Jourdenais 1998, Leow 1997, White 1995 & 1998, Lightbown, Halter & White 2002 and Izumi 2002).

1.5.4 Problems with research into enriched input

Clearly these enriched input studies have lead to more refined questions but definitive conclusions are not yet available. There are several concerns with this strand of research.

- i) Ellis (1999) and Izumi (2002) point out that there is a relatively small number of studies and that many involved short-term and limited amount of exposure to the target features.
- ii) Studies need to provide evidence for 'noticing' of the target forms (difficult if this is linguistic-context dependent) and for their subsequent incorporation into the interlanguage. Alanen (1995) suggested that noticing seemed to be induced by a variety of factors, not just input enhancement. Similarly, White (1998) reported that many learners noticed the forms but were not sure of their relevance or importance, which, arguably, accounted for the limited improvement by the enhancement group in her study. Jourdenais et al. (1995) did not assess IL development, but assumed a positive correlation between the 'noticing' they demonstrated, using data from think aloud protocols, and subsequent SLA.

- iii) Some studies suggest that visual input enhancement may work best in combination with other approaches such as: 'semantic elaboration' (e.g. comprehension assistance for each sentence, Doughty 1991), a form-focused recall production task (Williams 1999), activation of prior knowledge (Shook 1994) or explicit instruction (Ellis 1993).
- iv) The techniques and underlying theories adopted by the studies above do not attempt to offer linguistic or psycholinguistic accounts of SLA processing. Izumi 2002 argues for the need to consider levels and types of processing in order to account for how "sensory detection can lead to learning" (p542). In agreement with Lee (1999), VanPatten (2000) and Lightbown (2001), there appears to be a paucity of studies and theoretical explanations of what form-meaning connections learners make during on-line comprehension (i.e. listening and reading in real time) and why. One pedagogical 'package' that claims to be based on theories regarding these connections is Processing Instruction (PI, VanPatten 1996, 2002 & 2004) and is presented in the next section.

1.6 A brief summary of Processing Instruction and its potential suitability for developing year 9 learners' French verb inflections

This section provides a brief description of PI in order to highlight some of the potential benefits of investigating it in the context of UK secondary school MFL classrooms. A tabular summary of PI studies can be found in appendix 4 and examples of PI materials from published studies are in appendix 5. Full details and a critique of PI and its theoretical underpinnings (Input Processing) are given in chapter 2, including a discussion of some of the terms used here (e.g. processing, form-meaning connections, attention, detection).

1.6.1 What is Processing Instruction (PI)?

VanPatten (VanPatten & Cadierno 1993, VanPatten 1996, 2002 & 2004) claims that PI has several distinguishing features which set it apart from the input-based approaches described above. He suggests that purely comprehension-based approaches, raising learners'

awareness or visual enhancement³⁰ are not sufficient for second language learning, as learners must be *forced* to process form-meaning connections, in particular for those forms which are redundant for communicative purposes. VanPatten's theory of Input Processing (upon which PI is claimed to be based, presented and discussed in detail in chapter 2), suggests some hypotheses regarding what learners attend to in the input and why. These are referred to as input processing strategies. They suggest that learners do not always attend to the input in ways that would be best for learning certain communicatively non-salient features³¹. In practice these hypotheses are operationalised as a series of teaching activities (PI) which encourage learners *not* to use their routine strategies and promote 'more effective' input processing strategies by structuring the input and requiring learners to perform certain tasks with this input.

VanPatten (2004) writes:

"What is critically different about PI from other treatments that have an input-orientation (e.g., textual enhancement, recasts, input flood) is that PI first identifies a potentially problematic processing strategy from the model of input processing ... and then provides activities that push learners away from that strategy. In other words, PI does not just determine what is a problem form or structure but why it is a problem vis à vis one of the learning mechanisms involved in SLA".³²

There are three main components to the current form of PI. First, learners are presented very briefly with a written explanation of the target language feature including: its description, a brief warning regarding what learners tend to do 'wrong' and why (i.e. why that feature is difficult to process in the input). This explanation constitutes a kind of 'advance organiser' (Terrell 1991) to help learners attend to the target feature in the subsequent input activities. Following this, learners complete a few sentence level controlled exercises (called 'referential activities') where learners are required to process aural and written input where interpretation of the target form is *essential* for task completion. These exercises are designed to provide learners with many opportunities to make correct form-meaning connections for the target form in the input. Then learners

³⁰ Nevertheless, Wong (2002), Benati (2001) and VanPatten (1996, p72) show visual enhancement of the target form.

³¹ Briefly, this refers to language which has little or no apparent propositional content. The notion is discussed in detail in chapter 2.

³² References to VanPatten (2004), and chapters in it, do not have page numbers as the book was only published in February 2004. References are to drafts sent to me by VanPatten, Collentine and Benati.

complete one or two activities which involve tasks relevant to the learners' world / opinions (called 'affective activities'). They provide learners with a structured input flood of the target form where learners may see / hear the form in use³³.

1.6.2 *The role of input in focussing on form to affect the output*

The model in figure 1.4 (always shown in previous PI studies³⁴) suggests that by altering input we can affect output. Although the model (and the conclusions in studies based on it) is over-simplistic, it is presented here to indicate that VanPatten proposes a pedagogical package based on the basic idea that it may be possible to affect learners' productions by intentionally altering how they process input. (The appropriateness of this model is discussed in section 2.2.4. It should be emphasised that VanPatten acknowledges many essential roles for output practice in SLA e.g. VanPatten & Cadierno 1993, VanPatten 1996, 2002 & 2003³⁵).

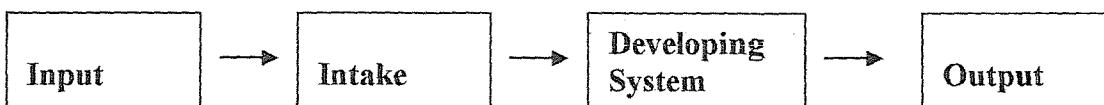


Figure 1.4 *Model of language learning often used in Input Processing literature.*

Most learners after a few hundred hours of exposure to the language (i.e. late Key Stage 3 / early Key Stage 4) are probably developmentally ready³⁶ to attend to verb inflections in the input, as verb inflections are already becoming productively emergent amongst many learners (see appendix 6)³⁷. For those learners whose verb inflections are *not* yet

³³ where processing the target form may not be entirely essential to the task (this is not addressed by VanPatten and colleagues but is argued in chapter 2).

³⁴ E.g. Allen (2000), Benati (2001). Such studies compare an input-based mode of learning with an output-based mode and as such, feed into debates regarding the extent to which language learning is modular (Fodor 1983) or a single complex system. They claim to find evidence against a skill-based model of SLA by demonstrating that input-based instruction can improve comprehension *and* production, rather than just comprehension, whereas production practice resulted in improvements in output only (see also Lightbown 1992). The current study does not aim to inform this debate, as discussed later.

³⁵ Though neither VanPatten nor his critics have discussed the role of output (target and non-target-like) as auto-input ('virtual input' or 'backdoor learning', Truscott 1998).

³⁶ Doughty & Williams (1998) suggest that it must be ascertained "whether or not learners have begun attempting the form in output or show evidence of comprehending it in the input" (p254) before an intervention can be effective (see also Spada & Lightbown 1999, Izumi 2002, VanPatten 1996).

³⁷ Macrory & Stone (1996 & 2000), Page (1999) and data collected for Mitchell & Dickson (1997) and for the Linguistic Development project (Myles 2002 <http://www.regard.ac.uk>) suggest that during semi-spontaneous oral production a few year 9 learners are still at the preverbal stage, many use apparently uninflected verb forms and some produce verb inflections (in target-like or non-target like ways). Learners' verb inflections

productively emergent, possibly because they are not 'noticing' the target form in the input, VanPatten (1990, amongst others) recommends activities that require learners to process them in input only.

1.6.3 Difficulties for L2 learners when processing the mature L2 verb inflection system

As argued above, the processing of input may be central in subsequent acquisition³⁸. The following section (using a range of evidence) suggests some of the reasons why learners may find difficulties in parsing the mature system of L2 verb inflection. This contributes to the argument that PI may be particularly suited to this language learning problem area.

a) Redundancy

Schumann 1987 (in Bardovi-Harlig 2000) points out that "in standard language, verb morphology interacts with, supports, and often duplicates work done by pragmatic devices" (p38). As French has an obligatory subject, marked for person and number, this could reduce the semantic importance of the verb inflections (except oral 3rd person subject pronouns before consonants, which are not marked for number - il / ils mangeØ – but, still, the verb inflection does not communicate any *more* information than the pronoun). The late emergence of verb inflections in learners' semi-spontaneous output, described in appendix 6, perhaps also reflects the redundancy of verb morphology during input processing.

b) Written and spoken systems

French oral and written input provides at best complex, at worst contradictory, evidence for learners (e.g. in spoken language there are three present tense *er* forms, in written there are five). This is partly due to the lack of transparency in the system of phoneme / grapheme

tended to be more target-like in written production. I could not locate any studies that have investigated the extent to which verb inflections are 'noticed' or comprehended by UK classroom learners (e.g. Harley 1992, in Canadian immersion programmes, used oral translation of verb morphology). This study makes some assessment of learners' interpretation of verb inflections.

³⁸ See also Harley (1994): "what [learners] take in can be seen to reflect the conditions in noticeability that have been specified. The children's productive repertoire in French has been found to consist mainly of phonologically salient, high frequency lexical items, along with syntactic patterns that are generally congruent with the L1...less salient morphosyntactic features of the target system, incongruent with the L1 and or not crucial for comprehension...may fail to become intake " p62.

relations (e.g. 's' is pronounced in certain linguistic contexts only, [e] can be represented graphically by ez, er and é) (L'Huillier 1999).

c) *Complexity of French verb inflections, in particular for Anglophone learners*

This section provides further evidence that input processing the French inflectional verb system can be problematic by briefly looking at the issue of complexity using the three theoretical perspectives (functional, formal and processing) proposed by DeKeyser (1998). These three perspectives are not related to the theoretical framework of the study (which is more embodied in the discussion of redundancy in a) but adds to a general impression of difficulty, particularly for Anglophone learners.

i) *Functional complexity*

Slobin (1979), Andersen & Shirai (1994), DeKeyser (1998) and Harley (1989) suggest that where one form is used for more than one meaning such forms are more difficult to acquire. The French present tense carries progressive and non-progressive meanings; the perfect tense is used for present perfective and simple pastness; [e] can indicate second person plural, 2nd person formal and some non-finite forms (including past participle); an oral null ending in French present tense refers to person (first, second or third), number (singular or plural in the 3rd person - the difference only audible before a verb beginning with a vowel), tense (present) and aspect (simple or progressive)³⁹. In addition, opportunities for learners to experience the most recurrent patterns of form / meaning mappings in the input may be reduced as highly frequent verbs tend to be irregular.

The fact that many inflections may *appear* homophonic to the L2 learner perhaps also misleads them into assigning more than one meaning to one 'apparent' form (Harley 1989 & 1992). Baddeley, Gathercole, & Papagno (1998) demonstrated that homophonic characteristics reduce the recall of new sounds / words. Given that it is unlikely that early Anglophone learners can distinguish rapidly, if at all, between [ɛ] and [e], input such as il rentr[ɛ] could provide positive evidence for learners that a form of the past with no

³⁹ Although this lack of one-to-one mapping is also present to some extent in the L1 for Anglophone learners of French, the complexity in L1 and L2 are different.

auxiliary does indeed exist (especially as about 90% of French verbs belong to the 1st conjugation (Clarke 1985)). Harley (1989) found that early immersion learners with 1000 hours of exposure to French used [ʒe] as part of the paradigm ‘être’: [ʒe], tu [e], il [e] (although it is acknowledged that this is output evidence).

ii) Formal complexity

French verb inflections could be seen as formally complex⁴⁰ as, for example, plurality must be marked on the verb in a variety of ways (e.g. 'ons', 'ez' or 'null' in oral regular 'er' present tense verbs).

iii) Processing complexity

In order for learners to make verbs agree with the subject, they must retain the subject in their working memory until they need to realise the appropriate inflection on the verb ending i.e. they must coindex the verb to the subject across at least the verb stem.

DeKeyser (1998) and Ellis (1990) suggest that the distance between the co-occurring elements can affect the difficulty of processing. DeKeyser (1998) and Pienemann (1998) offer subject verb agreement as an example as elements such as the lexical verb, noun phrase complements and / or clitic particles ('ne', object pronouns) can increase the distance between the grammatical number / person of the subject and the verb agreement⁴¹.

1.6.4 Growing research tradition & what other researchers say about PI

Another important reason for choosing PI was that research focussed on PI has a systematic line of investigation (see VanPatten 2002 & 2003, and appendix 4). Several PI studies meet the methodological criteria applied in Norris & Ortega (2000) to merit inclusion in their meta-analysis and aspects of the methodology used in VanPatten & Cadierno (1993a & b) are also praised by DeKeyser (1998) (p61). The PI tradition also attracts considerable debate amongst researchers (e.g. DeKeyser et al. 2002). The following is a brief review of

⁴⁰ Where formal 'simplicity' would be the presence or absence of a single morpheme communicating one 'meaning'.

⁴¹ Although the first two elements can also 'intervene' in English, this issue may pertain regardless of learners' L1. VanPatten 2003 suggests that heavy demands on working memory to co-index features is why learners do not acquire subjunctive verb inflections easily in Spanish. See also (Hawkins 1989) and (Myles 1995).

several commentators who, directly or indirectly, suggest that PI merits further investigation.

Doughty (2001) endorses PI as one of the few psycholinguistically motivated instructional packages available: "whereas pedagogically oriented discussions of issues ...abound, psycholinguistically motivated rationales for pedagogical recommendations are still rare" (p206).

Doughty & Williams (1998) see 'task essential' language as a positive characteristic as it renders the FonF unobtrusive, and illustrate this with a PI activity (p211). They point out that such tasks are easier to design for comprehension activities than production activities (see also Leeman et al. 1995).

Many applied researchers agree that FonF should constitute a combination of techniques and Doughty & Williams (1998) present a combination which could describe PI very well (p243). Lightbown (1998) writes "when explicit grammar teaching is done... VanPatten's 'input-processing instruction' [sic] is a good model of the type of grammar teaching that these separate lessons or minilessons might take" (p194).

Truscott's (1998) highly critical review of FonF studies associated with the Noticing Hypothesis (Schmidt 1990 & 1993⁴²) also does not include any PI studies. This absence is considered significant⁴³: Truscott firmly supports the notion that affecting the underlying system must be done via input and so, indirectly, endorses the PI approach.

1.6.5 '*Usefulness*'

Ellis (1995) suggests '*usefulness*' is an important criterion for evaluating a theory:

"the whole question of application [of a theory to a pedagogical context] cannot be addressed from the perspective of the theory to be applied; it must take as its

⁴² Briefly, the claim that second language learners must consciously notice the grammatical form of their input in order to acquire grammar.

⁴³ Truscott's does not critique the PI agenda even though it had at that time produced little evidence that PI affects spontaneous oral production (one of Truscott's main criteria for 'good FonF' studies). Nor can this be due to simple oversight of VanPatten's work: references *are* made to VanPatten's reservations about the direct impact of metalinguistic instruction on competence and his suggestion that input should be structured.

starting point the needs of language teachers. If there is a gross mismatch between the theory and the teachers' needs, application becomes difficult and perhaps impossible, no matter how 'good' the theory is considered to be" p87

One of the attractive aspects of VanPatten's work is that practical implications for the classroom are suggested. Two issues are suggested here that indicate that PI may be particularly suited to UK Key Stage 3 French classrooms.

a) Teacher control Vs learner autonomy

Teachers often express their desire to maintain tight control over learners' activities (e.g. Page 1992, Mitchell & Hogg 2001), whether researchers agree or disagree. Amongst other reasons (e.g behavioural and motivation issues), these reservations appear to stem from the impression that learners, left to their own devices, focus their attention on issues that are not intended by the classroom teacher (as expressed in some of the post observation teacher interviews in the action research projects discussed in section 1.3). This is corroborated by empirical studies: learners don't *naturally* seem to focus on grammar (Swain 1993, Williams 1999), and even less on the grammar intended by the teacher. PI provides a structured environment forcing learners to attend to the target features in ways that seem likely to appeal to MFL teachers in the UK, who are also often faced with motivation-related discipline problems (for example, listening and reading activities are often used as 'calming activities'). Another reason why PI's listening and reading activities may be welcomed is that plenty of anecdotal evidence suggests that productive skills are regarded as the most difficult by teachers and learners.

b) Time Efficiency

The limited time allocated to MFL learning in the UK makes the claims of PI extremely appealing i.e. that given the same amount of time and the same number of practice items as an output practice group, PI learners can subsequently perform *equally* well on production tasks and outperform the comparison group in their ability to interpret input.

1.7 Summary of chapter 1

It was suggested that the extent to which MFL grammar is taught in UK secondary schools is variable, and a great deal of teaching focuses on lexical items and functional phrases. When grammar teaching does occur learners' attention is usually drawn to grammatical features by explicit presentation (arrived at inductively or deductively) and automatisation of this knowledge is done via grammar-translation and/or behaviourist style activities. Given that the relationship of explicit knowledge with ability to use the language is not clear, that current policy documents used in schools can suggest few concrete recommendations informed by SLA research, and that there is dissatisfaction felt regarding learners' ability to manipulate verbs accurately, additional ways of focussing learners' attention on form are desirable. Several points raised also indicated why year 9 learners of French were chosen for this study: learning MFLs is obligatory at this age, requiring grammar pedagogy which is suitable not just for academically oriented pupils; dissatisfaction is often most acutely perceived towards the end of key stage 3 (i.e. year 9); process studies suggest that some learners at this stage are beginning to use inflectional verb morphology and that most are likely to be developmentally ready to benefit from noticing these forms in the input. It was argued that PI stands out as being potentially suitable for this context for several reasons:

- it aims to encourage learners to attend to communicatively redundant linguistic features that learners can 'ignore' in most current listening and reading activities;
- French verb inflections are complex to process in the input, suggesting that an input-based approach may be beneficial;
- it is based on a psycholinguistically-motivated set of principles;
- PI has a growing research tradition which so far suggests positive benefits for the package;
- PI may offer an environment compatible with many UK MFL teachers' views on classroom management and grammar pedagogy.

Chapter 2 The theoretical framework behind the pedagogical packages: "Processing Instruction" and "Enriched Input"

Introduction

The previous chapter suggested, through a review of educational literature, materials and classroom practice, that adapting what learners are asked to do with input may offer effective grammar pedagogy techniques for the learning of inflectional verb morphology amongst early learners of L2 French. Processing Instruction (PI) was introduced as a promising technique to investigate this.

This chapter presents PI in detail (section 2.1) and the theoretical framework that is claimed to underpin it, Input Processing (IP) (section 2.2). A range of concerns with the theory is included, although a comprehensive critique of IP (for example, its compatibility with other frameworks) is beyond the scope of this thesis. Section 2.3 examines the extent to which PI is driven by IP, discussing each aspect of the pedagogical package in terms of the theory and empirical evidence reported in the PI literature to support it and concluding by suggesting that PI is a reasonably faithful operationalisation of IP theory. In section 2.4 the comparison made in this study between PI and Enriched Input (EI) is explained. The research questions, a summary of the chapter and the unique features of this study are presented at the end.

2.1 Processing Instruction

The four main components of PI are explicit grammar instruction, referential activities, affective activities and some guidelines. These are each simply described in this section, as presented by the proponents of PI. Examples of materials from VanPatten & Wong (2001), PI for the French 'faire causatif' for university students, are given in appendix 5. Terms such as 'attention', 'notice', 'process' and 'form' are defined and discussed in more detail in sections 2.3 onwards. Until then, it is acknowledged that they require a broad, relatively superficial interpretation.

2.1.1 *Explicit Grammar Instruction (EGI)*

EGI here is meant as “explanation about properties of language provided by an instructor, teaching materials or some other external sources” (VanPatten & Oikkenon 1996, p. 6).

The EGI in PI consists of a brief presentation of the grammar point in focus *before* the learners are given practice opportunities. First, the target feature is explained in terms of the form-meaning connection learners must make. Then there is a statement about the type of errors that are normally made by learners, the reason why learners make such errors and one or two examples of them. This is in order to "inform learners of the potentially problematic outcomes of processing strategies" VanPatten 1996 (p62) ('processing strategies'¹ are presented in section 2.2). PI requires that the EGI is kept as brief and clear as possible and that whole paradigms are not presented at once.

The EGI should not be kept available for the learners once the practice element has begun, although learners may be reminded of the problems of processing input at the start of referential or affective activities e.g. "Keep in mind that Spanish does not follow a rigid subject-verb-object word order and that object pronouns may go before...." (VanPatten 1996 p63). Also, other uses of a type of EGI are apparent in *other* parts of the PI package. For example, the following short extract is from an affective activity: "find all the uses of a third person object pronoun. To whom or what do they correspond? What is the subject of each verb next to which you found each pronoun?" (VanPatten 1996 p78).

2.1.2 *Referential Activities*

The first practice activities are at the word / sentence level and focus learners' attention on specific features of the language by requiring them to interpret the meaning of those features in order to carry out the task set. There is a right or wrong answer. The target feature is placed in utterance initial position, or as near as possible. Learners are guided by feedback from the tutor after each of a few items at the start of an activity to ensure they are making the correct form-meaning connections. VanPatten (2002) suggests that "normally,

¹ e.g. learners not attending to the verb inflection 'ons' because its meaning is carried in the pronoun 'nous'

a sequence of structured input² activities would begin with two or three referential activities" p766.

A crucial defining feature of referential activities is that the target feature is compared explicitly with another, usually similar feature, which may be the source of some confusion for the learners or may help to emphasise the distinctiveness of the target feature. This characteristic of referential activities was only implicit in the PI literature at first, but it seems to be gradually taking on greater importance, and this is discussed later.

2.1.3 Affective Activities

Following referential activities, 'affective activities' use one or more exemplars of the structure in the context of some 'real world' task. Learners must express an opinion, belief, or give some other affective response about phenomena associated with their tutor, their own lives, well-known personalities, a joke or cartoon etc. There are no right or wrong answers during these activities. Though not acknowledged elsewhere, the tasks do not usually *require* the learners to process the meaning of the target features in order to complete the task i.e. any attention they give to the target feature is usually incidental to the task.

However, a further characteristic of affective activities is that there is often a 'reminder' to the learner to pay attention to the target form e.g. "Notice that the verbs are all expressed in the future tense (third person)" (Benati 2001 p123). This can sometimes include a question which does force some attention to the meaning or function of the target feature e.g. "Have you noticed the future forms? Are they first or second person forms? How do you know?" (Benati 2001 p124).

2.1.4 Guidelines for Processing Instruction

VanPatten's guidelines are simply listed here as in Lee & VanPatten (1995) and VanPatten (1996). VanPatten (1996) emphasises that they should be used flexibly according to the learning group (p67). They are discussed in greater detail in section 2.5.

² Referential and affective activities combined are known as 'structured input'.

- “1 Teach only one thing at a time
- 2 Keep meaning in focus
- 3 Learners must do something with the input
- 4 Use both oral and written input
- 5 Move from sentences to connected discourse
- 6 Keep the psycholinguistic processing mechanisms in mind”.

VanPatten 1996 (p67)

2.1.5 *Previous PI studies: Some design problems*

As mentioned in chapter 1, there is a substantial and growing tradition of research into this pedagogical package (see appendix 4³ for a tabular summary; Ellis 1999 for a summary and critical evaluation, and VanPatten 1996, 2000, 2002 & 2003 for summaries). Most studies have involved university learners and Romance languages, with two main linguistic foci (syntax of object pronouns and a range of inflectional verb morphology). Intervention treatments have been fairly short (often about a couple of hours). Most post tests have been immediately after the intervention and delayed post tests after approximately three weeks. Measures of interlanguage development have generally been controlled sentence level tests, usually aural interpretation and written production tests. Such measures have been subject to criticism from SLA researchers, because they may tap into explicit knowledge rather than the underlying IL system. Better indicators of changes in the developmental system according to most SLA researchers (e.g. Norris & Ortega 2000, Truscott 1998) include: grammaticality judgement tests, long term post tests and semi-spontaneous oral production. Some studies have begun to address some of these criteria VanPatten & Sanz (1995)⁴ and VanPatten (2004).

Notwithstanding the concerns outlined above, most of the studies have found that when compared with some form of output practice PI has favourable effects in terms of learners' interpretation of the target features and equivalent or favourable effects in terms of

³ Some of the unpublished studies reviewed in Ellis (1999) have been included in this table, even though they have not been reviewed first hand and VanPatten considers that they do not constitute replications of PI (personal communication, 14 June 2002), despite some support for PI therein.

⁴ VanPatten & Sanz (1995) almost found statistically significant improvement on an oral narration task (though there was no comparison group, only a control with 'no' instruction). It is acknowledged (here and by VanPatten) that even measuring 'acquisition' in this way is not without controversy, as speech processing constraints are likely to be a highly influential factor. No published study has yet attempted to make any direct measurement of 'intake' (e.g. by using implicit memory tests or grammaticality judgement tests) even though VanPatten and colleagues suggest PI improves 'intake' (discussed later).

learners' production of the target features. Favourable effects in both interpretation and production have also been found when compared with a control group with no instruction in the target feature.

A detailed review of the sustained research agenda involving PI is not provided here because of a central design problem which affects these studies and makes them not directly comparable with the present study. It has just been noted that the comparison instruction in previous PI studies is output practice. It is therefore suggested here that no study has yet satisfactorily explored the central claims of IP theory regarding what learners do with the input and why. By nature of their comparison groups these studies are not able to claim that it was the precise *nature* of the input and input-based tasks that led to effective learning by guiding learners away from the processing strategies suggested in IP (discussed later). For example, the studies do not address the fact that the apparent benefits of PI could be explained only by the different practice modalities i.e. input rather than output practice. In this sense, previous studies have addressed a fundamentally different question from the current study i.e. the modularity and skill-based nature of SLA, as seen in the discussions in DeKeyser (1998), DeKeyser & Sokalski (1996) and Ellis (1999).

2.2 Input Processing theory

IP theory, its evidence and related discussions were originally laid out in VanPatten & Cadierno (1993), presented more fully in VanPatten (1996) and revised slightly in VanPatten (2000) and VanPatten (2004). IP theory consists of a body of writing about the processes involved with input processing in SLA and a set of principles which largely summarise this writing by suggesting a list of strategies that learners use when decoding input. Two attempts have been made to represent a model of IP graphically and these are given in appendix 7 (a & b). It is emphasised that the purpose of this study is not to scrutinise these models (particularly as it is considered they do not add anything substantial to IP).

This section first presents key notions central to IP, which generally emerge from cognitive psychology. IP's conceptualisation of 'attention' is discussed next, as this has particular

prominence in IP theory. The principles and a summary of the evidence offered for them are then presented. Several important challenges to IP are summarised to acknowledge the limits of IP, though these cannot be significantly addressed in this study. However, one challenge which may be relevant to this study is then discussed: the role of the L2 learner's developmental system in input processing. The section ends with a summary of the theoretical notions central to this study.

2.2.1 *Key Notions in IP*

This section presents key theoretical constructs originating from cognitive psychology that have been adopted and developed by IP theory. In order to select relevant work, I have followed many conceptual links from VanPatten's work and from other SLA publications e.g. Ellis (1999), Segalowitz & Lightbown (1999), Schmidt (1993), Schmidt (2001), Simard & Wong (2001). The complex notion of 'attention' and its conceptualisation within IP is discussed in 2.2.2.

a) "Input Processing"

Input processing can be viewed from a variety of perspectives. Considerable reference made by VanPatten to concepts from cognitive psychology situates IP within a broadly cognitive view of SLA. VanPatten (1996) writes "The job of input processing is to detect linguistic data and make initial form-meaning mappings, even if they are incomplete or partial... This input processing provides the intake for further cognitive processing" (p31). VanPatten (2004) suggests: "processing refers to making a connection between form and meaning. That is, a learner notes a form and at the same time determines its meaning (or function). The connection to meaning may be partial or it may be complete". VanPatten 1996 defines processing as "attending to and detecting linguistic data in the input" (p17). Clearly, the terms 'attend' and 'detect' are therefore crucial to the interpretation of IP.

b) "Detection"

Central to IP is that 'detection' of an input feature, a sub-process of attention, is essential if it is to be maintained in working memory for further processing. VanPatten adopts the definition of detection proposed by Tomlin & Villa (1994) i.e. "the process that selects, or engages, a particular and specific bit of information" (p192). VanPatten specifies that this

involves not just perceiving a form but *connecting a meaning to it*. VanPatten claims "detected information causes great interference with the processing of other information and the detected information exhausts more attentional resources than even orientation of attention" (VanPatten 1996 p17). The role assigned to detection is what principally distinguishes IP from Schmidt's Noticing Hypothesis (Schmidt 1990 & 1994). Both agree that attention to form in the input is essential for learning to occur. However, VanPatten 2004 points out the distinction between noticing and detecting a form:

"Noticing...refers to any conscious registration of a form, but not necessarily with any meaning attached to it (Schmidt, 1990)...perception and noticing ...do not necessarily imply that a form has been processed (linked with meaning...)."

c) *"Form-meaning connection"*

VanPatten (1996) defines form-meaning connections as those "that the learners' internal processors make between referential real-world meaning and how that meaning is encoded linguistically....Form-meaning mappings happen during input processing and are necessary for the building of mental representations" (p10). Note that VanPatten's definition does not include the *production* of a correct form, unlike Terrell (1991), as IP is concerned with initial form-meaning connections only, which may or may not, correctly or incorrectly, make their way into the developing system.

d) *"Intake"*

According to VanPatten (2004), intake "refers to that subset of the input that has been processed in working memory and made available for further processing (i.e. possible incorporation into the developing system)". VanPatten (2000) argues that the processes involved in deriving intake from input occur during the act of comprehension. The notion of a 'reduced' or 'filtered' input due to online processing constraints during comprehension has been expressed by others (e.g. Pinker 1982 p667, Frazier & de Villiers 1990 p10-11, Tomlin and Villa 1994, Lightbown 2001). However, VanPatten (2004) specifies, "intake is not just filtered data (i.e. a mere subset of the input) but it may include data processed incorrectly (i.e. the wrong form-meaning connection may be made)". VanPatten (1996) suggests that this reduced and sometimes altered subset of input means that the developing system does not receive all the data needed to instantiate aspects of the internal grammar

(VanPatten suggests that it is UG that acts on intake data in this developing system p134⁵).

This supports the proposal that L2 input processing may deprive and misinform the developing system of the information it needs for the accurate construction of the L2.

VanPatten points out that Towell & Hawkins (1994) also suggest that processing limitations within working memory may mean that input data cannot correctly confirm or reject the internally derived hypotheses "because short-term memory is limited in capacity [and] learners are often unable to decipher the internal structure of utterances when these are first heard in context" (Towell & Hawkins 1994 p253).

2.2.2 "Attention"

VanPatten (1997) suggests that attention to form in SLA refers to "what learners hold in working memory long enough so that connections between form and meaning can be made [i.e. detection] at the sentence level (cf. Tomlin & Villa 1994)" (p94). According to IP 'attention' is therefore essential if detection (and therefore learning) of forms is to take place. A brief overview of how attention has been perceived in cognitive psychology and how it is conceptualised within IP is given here. The role of 'attention' in learning was rejected by leading proponents of classical behaviourism (Neumann & Sanders 1996) and of innatist language acquisition theories but is now becoming a central construct in many SLA theories (Schmidt 2001, N. Ellis 1999)⁶. However, application of the construct is proving to be problematic as it can be poorly defined (as noted by Schmidt 1993, VanPatten 1996, Simard & Wong 2001, Truscott 1998, Harley 1994 and Anderson 1995). This section aims to clarify how 'attention' is conceptualised within IP.

a) Overview of cognitive psychology research on attention

IP's view of attention conforms to much of the mainstream cognitive psychology literature regarding the definition of attention (Schmidt 2001, Miyake & Shah 1999). Of the six characteristics from cognitive psychology literature outlined by Schmidt (2001), VanPatten

⁵ However, VanPatten 1996 emphasises that "*input processing is not an account of learning; it is an account of what kind of intake data are made available for learning*" p147 (emphasis original) and it is therefore suggested that regardless of one's perspective of the developing system, this reduced and incorrect intake may impair its functioning. E.g. models of attention in the 1980's began to incorporate connectionist theories in which the effect of attention is conceptualised as an additional input into the units representing the selected stimuli (Cohen, Dunbar, & McClelland 1990)

⁶ or, at least, SLA theorists acknowledge their model does not account for attentional processes (e.g. Pienemann 1998 "(Processability Theory) is not "designed to contribute anything to the question of ... the inferential processes by which linguistic input is converted into linguistic knowledge" p5.

clearly endorses five: attention is limited, selective, partially subject to voluntary control and is essential for learning and for the control of action (i.e. the earlier stages of automatisation). The remaining 'classic view' of attention is that it controls access to consciousness, but VanPatten suggests that attention and detection may be with or without consciousness. The arguments surrounding this are not reviewed here as the notion of consciousness could not be pursued in this teaching intervention study, though it is acknowledged that this remains a controversial aspect of IP theory.

Within cognitive psychology there have been four main phases of research perspectives on attention since 1958, when the construct regained respect largely due to Broadbent (1958) who formulated a workable set of hypotheses and experimental paradigms (Neumann & Sanders 1996). VanPatten's IP shares many theoretical perspectives with phase one, the bottleneck filter model (Broadbent 1958). This model, although based largely on acoustic processing, also suggests that attention cannot simultaneously be directed at form and meaning (although the filter theory differs fundamentally from IP as it claims that selective attention operates on the form of the message first)⁷. The notion of attenuation i.e. graded, rather than 'all or nothing' selective attention (Broadbent 1971) is not, however, incorporated into IP. Phase two was characterised by the idea of controlled versus automatic processing (Posner & Snyder 1975 and Shiffrin & Schneider 1977) and this is also reflected in IP in the discussion of automaticity enabling, over time, the parallel processing of form and meaning. Phase three introduced the notion of multiple resources (Wickens 1984) and is seen in the Availability of Resources Principle discussed below. There has been criticism (DeKeyser et al. 2002) that VanPatten adopts an outdated model of attention because his work is not influenced by a fourth phase of attention research where the notion of a limited capacity has been abandoned. This is discussed next.

⁷ Another relevant aspect of the filter theory is the suggestion that "in monotonous conditions the filter becomes 'satiated' in selecting the same type of information over and over again" p2 (Neumann & Sanders 1996). This may well be an important factor which is yet to be formally explored within IP and PI e.g. whether learners become 'immune' to PI after a certain number of items / activities / sessions, as VanPatten commented during the design of my materials "the second activity is terribly long and might induce the boredom factor...My experience is that about 10 items is maximum before you induce boredom in any one activity" (Personal communication 13 November 2001).

b) Limited Capacity

VanPatten situates himself within a considerable body of studies which supports the concept of a limited capacity of attention whilst performing related functions. VanPatten (2002) claims that unlimited capacity models of attention are not intended to explain language processing. VanPatten cites Just & Carpenter (1993) (cognitive psychologists/ psycholinguists who developed the only language comprehension-based model of capacity limitation) to suggest that although successful simultaneous processing of competing stimuli can occur in adult native speakers, it cannot occur if the amount of activation exceeds a certain amount, in which case the processes propagating this activation will be scaled down. According to Just & Carpenter both storage and computational processes can be down-graded as both these are carried out in working memory. VanPatten also cites Carr & Curran (1994) who concluded that even though learning may seem to happen without awareness, when a concurrent task is introduced learners' performance deteriorates markedly or becomes impossible⁸. Wickens (1984) and Gathercole & Baddeley (1993) are also both used by VanPatten to support the notion that when information is confined to one modality (e.g. aural) then attentional resources to process these competing stimuli are particularly constrained by limits in capacity.

In support of the limited capacity model of attention is the fact that the concept is shared, in some form, by most cognitive psychologists (Miyake & Shah 1999), psycholinguists (Pinker 1982) and many SLA researchers (e.g. Towell & Hawkins 1994, Myles 1995, Schmidt 2001, Towell & Dewaele 2002).

It is emphasised that limited capacity models do allow attention to be allocated to more than one task but only if one or both can be performed automatically – see sections 2.2.3 and 2.2.5 for further discussion of limited attentional resources and the role of developmental system in input processing.

⁸ Concurrent tasks can involve noting occurrences of a particular item or verbalising something. It is acknowledged that such studies using artificial 'grammars' e.g. letter strings, do not necessarily replicate SLA, due to the nature of the tasks and the lack of meaning and linear structure of the grammars. They are cited here in line with most SLA attention research (this perhaps indicating the need for further research on simultaneous processing specific to learning natural second languages).

c) *Selective Attention*

As the capacity for attention within working memory is limited, learners must select what they attend to. VanPatten (1996) refers to Lachman, Lachman & Butterfield (1979) to support the concept of "selective attention bringing particular stimuli into focal attention rather than allowing them to be merely perceived" p15. Selective attention would predict that certain items will be 'suppressed' in order to leave sufficient resources to process other items, enabling humans to select relevant stimuli from the vast range available at any one time. Similarly, IP predicts that items of higher communicative value (discussed later) will be attended to and detected (attached to a meaning i.e. processed) whereas items of low communicative value will be suppressed as they constitute less or non-essential information.

Broadbent's (1958) filter theory of attention is also possibly relevant to VanPatten's conceptualisation of attention as it encompassed notions of 'selective listening' and 'automatic detection'⁹. The former suggested that the filter could be strategically set toward certain signal sources, presumably as intended by PI referential activities. The latter suggested that the filter could also be automatically attracted by certain signals, presumably the 'inefficient strategies' described by the Principles of IP, which are presented next.

2.2.3 *Input Processing Principles and their empirical and theoretical base*

The IP Principles claim to predict strategies¹⁰ learners are likely to use when selecting those forms from the input which will undergo further mental processing and so become intake¹¹. As such, they serve as a guide for choosing the linguistic focus and design of instructional material, as described here and in the methodology chapter.

Whenever possible, the application of IP Principles is illustrated with reference to the linguistic focus of this study (i.e. French verb inflections for tense, person and number).

⁹ The term detection was not used by Broadbent in the same way as VanPatten uses it.

¹⁰ The term 'strategy' is used as in cognitive psychology conceptual frameworks of learning and memory and refers to mental steps, mechanisms or operations carried out to accomplish cognitive tasks (e.g. memorisation, processing information). The term is not used synonymously with the "learning strategies" literature.

¹¹ VanPatten does discuss, albeit in a rather cursory manner, how items in the input are identified in the first place, referring to, for example, Peters 1985 Extraction and Segmentation principle, UG, perceptual saliency. Detailed discussion of these is beyond the scope of the thesis, though some issues relating to them are raised later.

This section presents Principle 1, its corollaries and evidence *as presented by VanPatten*¹². Further evidence or suggestions for research which may more fully inform the Principles are given, though as stated above, critical discussion is limited and challenges are summarised in section 2.2.4. Principle 2 (the first noun principle), regarding syntactic parsing, and its corollaries are not discussed in this thesis as they are irrelevant to this study¹³.

a) *Principle 1: The Primacy of Meaning Principle*

'Learners process input for meaning before they process it for form'

The overarching statement for Principle 1 is supported in VanPatten (1996) by appealing to similar observations made by other researchers in both first and second language acquisition: for example Peters (1985), Sharwood Smith (1986) and the 'Operating Principles' in Slobin (1979). A broad interpretation of the notion that form and meaning cause a tension in the learners' processing mechanisms is also shared by a wider body of researchers than PI studies refer to e.g.: semantic versus syntactic processing (Swain 1995); processing for communication and processing for acquisition (Sharwood Smith 1986); "semantic comprehension is a prerequisite for syntactic comprehension, and syntactic comprehension is a prerequisite to acquisition" (Gass 1997 p.137). McLaughlin & Heredia (1996) suggested that the top-down effort of processing input for meaning is such that learners are not able to notice the formal features of the language. Ranta (1998) and Segalowitz (2000) both suggest that there is considerable individual variability in learners' ability to change the focus of their attention from meaning to form. Lightbown (2001) uses evidence from immersion contexts and concludes that "learners treat some features of the input as transparent, because the meaning is clear; or as unimportant, because they do not appear to carry significant meaning" (p94).

However, it is acknowledged here that dichotomous use of the terms 'Form' and 'Meaning' probably constitutes a false dualism. Categorising some aspects of language as 'form' and others as 'meaning' is simplistic as every language feature has 'form' – be it phonological,

¹² It is noted that much of this evidence dates from the 1970s and 1980s, suggesting that PI theory originated from a pedagogical concern regarding the lack of full acquisition in L2, which was largely expressed through the traditions of error and contrastive analysis.

¹³ See appendix 8 for the complete set of IP Principles.

graphic or syntactic – regardless of the transparency of its link to semantic concepts.

Although VanPatten (1996) acknowledges that he does not intend to "pit form against meaning" (p47), it is suggested here that his definition of form ("verbal inflections, nominal inflections, particles, functors, and so forth", VanPatten 1996 p18) would benefit from refinement¹⁴. Without this, it is argued that the overarching statement "Learners process input for meaning before they process it for form" is somewhat cavalier.

Nevertheless, it is suggested that the corollaries to this principle (presented next) may inform pedagogical decisions regarding what language features to focus on, as they give some indication as to which may be more likely to be more communicatively useful to learners whilst comprehending. (Though some refinements could also be made to these corollaries, for example, there is a lack of clarity in what 'content words' and 'lexical items' refer to¹⁵, and more detailed discussion is required regarding their phonological and syntactic 'form' and how these are processed¹⁶).

b) Principle 1a: The Primacy of Content Words Principle

'Learners process content words in the input before anything else'

For empirical evidence for this sub-principle VanPatten draws on both observational studies in L1 (Peters 1985 and Radford 1990) and SLA (Klein 1986 and Mangubhai 1991) and experimental studies from SLA (VanPatten 1990). He notes that Peters (1985) and Radford (1990) showed that in child L1 acquisition the child focussed on isolated words and unanalyzed chunks and incorporated these into their production. Klein (1986) carried out an elicited imitation task showing that learners could not incorporate grammatical items such as auxiliary verbs and articles in their imitations of the investigator's speech¹⁷. Mangubhai

¹⁴ One illustration of how complex this issue is that Doughty & Williams 1998's definition includes "forms (e.g. phonemes, morphemes, lexical items, cohesive devices and politeness markers) and rules (e.g. devoicing, allomorphy, agreement, collocation, anaphora and in-group vs. out-group relationships)" p212.

¹⁵ VanPatten (2000) notes that 'content words' includes "lexicalized chunks of language or routines and patterns that are fixed and invariant for the learner" p 299, though this is not very helpful in clarifying the differences.

¹⁶ VanPatten (2003), somewhat crudely, suggests that L2 learners already carry, in their Universal Grammar, the notion of 'content words' and that the detection of lexical items in the first instance is helped by acoustic features of content words (usually receiving stronger stress). However, these issues are beyond the scope of this classroom based study which would not be able to inform why or how learners tend to initially process items in a sound stream.

¹⁷ Further evidence of this was found in the tests in this study.

(1991), using think-aloud protocols¹⁸, found that all learners reported that they focussed on 'lexical words' in order to get meaning from the input during a course of 'Total Physical Response'. VanPatten (1990) showed that learners were able to note occurrences of the noun 'inflación' and successfully perform free written recalls yet recall was increasingly hampered when asked to note definite articles or verbal inflections. If one assumes that more easily detected items represent less of a burden on limited capacity processing resources, leaving more room for the retention of information for the written recall, then this study constitutes significant evidence in favour of the notion that lexical items are processed more often. VanPatten (1990) provides a review of related work that supports principle 1a).

Additional evidence for this corollary could be found in the tracking literature, such as Bernhardt (1987) who followed eye movement of beginner L2 French learners and found that when they read they 'looked at' the nouns and the roots of verbs whereas the more advanced L2 readers revisited words to look at inflections, even more so than L1 readers. Such studies, as yet untapped by IP literature, may help to position IP within other perspective on L2 reading e.g. Lee (1992) reviews studies of reading amongst beginning level language learners and posits that beginning level language learners can and do engage in both bottom-up and top-down driven processes.

c) *Principles 1b, c & d and the notion of communicative value*

Together, corollaries 1b, c and d express the concept of Communicative Value (VanPatten 1985) which involves the notions of redundancy and inherent semantic value, as explained below.

Principle 1b: The Lexical Preference Principle 'Learners will tend to rely on lexical items as opposed to grammatical form to get meaning when both encode the same semantic information'.

¹⁸ This thesis acknowledges that think-aloud protocols do not necessarily reflect internal mechanisms and suggests that more sophisticated means are required to dismiss conclusively the existence of simultaneous implicit / unconscious processing of 'non-lexical' items.

Principle 1c: The Preference for Non-redundancy Principle 'Learners are more likely to process non-redundant meaningful grammatical form before they process redundant meaningful forms'.

Principle 1d: The Meaning-before-nonmeaning Principle 'Learners are more likely to process meaningful grammatical forms before nonmeaningful forms irrespective of redundancy'.

Structural features (e.g. position, syllabic length) cannot fully explain why certain features are acquired before others (e.g. in English plural 's' is acquired prior to singular 3rd person 's'; in Spanish verbal inflections -o,-n, -mos, -s, etc are acquired before adjective concordance -o, -a, -s, -ø, (van Naerssen 1981)). It is likely therefore that some factor other than structural characteristics must be involved. IP suggests that one of the major problems for L2 learners is that the referential meaning of functional or seemingly abstract language features is often shared by other features in the language e.g. the pastness communicated by verbal inflections is often also communicated by temporal adverbs; interrogative word order by WH words and/or intonation and/or verbal morphology. VanPatten and colleagues suggest that this phenomenon (known as Communicative Value, CV) affects the detection of features in the input.

CV is based on the 'inherent semantic value' and 'redundancy within the utterance' of the linguistic feature. Features with higher CV have higher inherent semantic value and are less redundant. Features of lower CV have lower inherent semantic value and are more redundant. VanPatten (2002) explains:

"A given form can have [+semantic value] and [-redundancy] (e.g., English *-ing*), [+semantic value] and [+redundancy] (e.g., subjunctive verb inflections), [-semantic value] and [+redundancy] (e.g., adjective concordance in Romance languages), and finally [-semantic value] and [-redundancy] (e.g., some complementizers such as that)... Forms with [-semantic value] regardless of redundancy contain no communicative value.... A form with no or consistently little communicative value is the least likely to get processed and, without help, may never get acquired" p4.

For example, in French overt subjects are obligatory and communicate both person and number lexically, whether nouns or pronouns¹⁹. Therefore verbal inflections for person and number, although also carrying inherent semantic value, have low CV both in writing and speech e.g. the '*ons*' ending has permanently low CV being redundant as it always accompanied by the subject. Furthermore, pronominal reference itself always has additional contextual support (both linguistic support via the higher CV of the noun that the 3rd person pronouns refers to, and other more pragmatic contextual factors e.g. who is talking to who, gestures etc). This may suggest that verb inflections are even less likely to be processed for person and number.

As acknowledged by VanPatten (1996), features do not have a fixed inherent semantic value nor fixed redundancy, as linguistic and other contextual features affect these properties e.g. an absence of temporal adverbs would decrease the redundancy of the morphological means of marking tense and aspect. However, it is common for temporal cues other than verb morphology to be present in the input, assigning a low CV to the verbal inflection for tense.

'Auxiliaries' and 'past participles' in the passé composé also raise interesting issues regarding the way learners may perceive their CV. In the passé composé the auxiliary carries inflection for tense. However, as the past participle is the lexical item with the propositional content, learners' attention (according to IP) may focus on this for meaning and, as the past participle is frequently distinguishable from other forms²⁰, they may interpret it as carrying tense (albeit incorrectly). This may be particularly likely if other uses of the past participle are not frequent in the input (e.g. *j'aurai mangé*). Aspects of formal instruction may promote this erroneous 'processing strategy': section 1.4.2b (appendix 1a & b) gave two examples of a teacher focusing learners' attention on the past

¹⁹ except oral 3rd person pronouns where there is no distinction between singular and plural forms when the following verb begins with a consonant or is preceded by *ne* / *n'* where there is no distinction between singular and plural forms.

²⁰ Orally, in regular *er* verb past participles, the verb ending [e] is distinct from all but one present tense inflection (2nd person plural) and infinitives, although also ending in [e], are always distinguished by their position after another lexical verb; the written 'é' could be misinterpreted as indicating tense. The oral form of regular *ir* verbs is distinguishable from present tense forms (e.g. *sorti*). Regular *re* verbs change to the very phonologically distinct 'u', dissimilar from any present tense. Some common irregular verbs *lu*, *vu*, *bu*, *pris* could all be misinterpreted to indicate tense.

participle to communicate the pastness of a sentence²¹. The fact that learners often use a past participle type verb in perfect tense contexts (e.g. no auxiliary but lu, bu, and written é Macrory & Stone 1996, Page 1999) possibly suggests that learners mis-interpret the CV of the past participle. A separate study would be required to investigate the prediction that the past participle may be more frequently processed by learners for carrying tense and whether this induces 'non-detection' of auxiliaries. In this study, particular use was made of constructions where only the auxiliary can be aurally interpreted for tense e.g. je fais / j'ai fait; vous regard[e] / vous avez regard[e]); j'aime jou[e] / j'ai jou[e] and just one 'unit' of teaching materials (see chapter 3) focussed learners' attention on the regular 'er' past participle.

More investigations are needed to refine the notions of CV, inherent semantic value and redundancy, as these are fluid characteristics and perceptual salience and frequency interplay with CV, as VanPatten (1996) acknowledges. Nevertheless, he offers a range of evidence for these corollaries as follows.

Findings that learners encode notions such as tense, plurality, person and number firstly by using lexical items are cited as evidence (e.g. Bardovi-Harlig 1992 and Pica 1985). Documentation of pidgins and creoles provides further support (Todd 1974)²², as do unpublished experimental studies such as Cadierno et al. (1991) and Musumeci (1989) and introspective studies e.g. Glass (1994)²³. Swain (1985) is cited as she found that in completely content-based environments learners' talk about language focuses overwhelmingly on lexical items, rather than morphosyntax (as in other studies, not cited by VanPatten: Williams 1999, Jones 1992 and Slimani 1991).

Studies more directly related to investigating the concept of CV are reviewed in VanPatten (1996). For example, Lee (1987) was a think-aloud study claiming to show how learners skip items of low CV during processing. Also, Bransdorfer (1989) carried out dual task

²¹ The use of terms such as *past* participle compared to 'helping' verb may also encourage this assumption.

²² Additional evidence for this corollary could be found in (Bardovi-Harlig 2000) includes a summary of the interlanguage phenomena of using lexical means to express tense before morphological means, including a summary of findings from cross-linguistic studies such as the European Science Foundation project in the 1970s (Perdue 1993)

²³ However, as discussed later, acoustic salience was not kept constant between the temporal adverbs and the verbal inflections and so could be an influential factor

attention experiments, measuring learners' written recall of idea units in a passage when they had to note occurrences of '*de*' (an item of high CV in Spanish, denoting possession) compared to noting '*la*' (a definite article of low CV which can usually remain undetected without confusing meaning). The findings suggested that learners found noting occurrences of '*la*' was much more costly to resources as they recalled significantly less information than the passage-only group, whereas the group noting '*de*' were able to recall a similar amount of idea units as the passage-only group. An important feature of this study is that some significant issues of perceptual salience were taken into account (both items were syllabic and occur before nouns) and as such it suggests a possible research agenda for Principles 1b, c and d.

Relevant to the context of this current study are the findings of an action research project in which 11 year old learners of German read sentences containing incorrect written subject + verb combinations and were engaged in retrospective think-aloud protocols (Mitchell & Hogg 2001). Prior to instruction there was a significant tendency for learners not to notice these errors but to suggest that lexical items were incorrect. Most relevant as evidence for the concept of CV is that when learners did notice a mismatch between the subject and verb all learners always chose to alter the endings to suit the subject, suggesting that the subject dictates to the learners what the required ending is, assigning the verb ending inferior status in terms of its CV.

d) Principle 1e: The Availability of Resources Principle

'For learners to process either redundant meaningful grammatical forms or nonmeaningful forms, the processing of overall sentential meaning must not drain available processing resources'.

This sub-principle rests on limited capacity attention models²⁴. It also links closely to the discussion of the role of learners' developmental stage in input processing in section 2.2.4. The evidence for this principle cited by VanPatten and some additional evidence is given in this section.

²⁴ It is not understood why VanPatten (1996) wrote: "There is as of this time no solid experimental evidence that directly supports this principle" (p27), as, although not undisputed, it would find considerable theoretical and empirical support amongst cognitive psychologists (Miyake & Shah 1999, Neumann & Sanders 1996).

VanPatten (1996 & 2003) refers to three SLA studies which, although methodological problems render them suggestive rather than conclusive, have the potential to inform and refine Principle 1e: Leow (1993), Berne (1989) and Blau (1990). Leow (1993) and Berne (1989) used simplified input (written and oral respectively) to operationalise the notion of reducing demands on finite resources. Leow found this had a favourable effect on learning to use the Spanish present perfect and subjunctive. However, Berne found that simplifying the input did *not* improve learners' ability to perform recalls whilst also noting the occurrence of a definite article or a verb inflection. VanPatten (1996) suggested that this may indicate that for learners to release attention to form, without adversely affecting their comprehension, the input may "require a *great deal* of simplification". Blau (1990) found that pausing had a more facilitative effect on the detection of less or non meaningful features than reducing the rate and syntactic complexity of input. VanPatten (2004) also argues that interactionist and input studies (e.g. Long 1985, Hatch 1983) show that repeating input (during negotiations of meaning) or simplifying input releases attentional resources so that learners are more likely to acquire / notice aspects of form. Other 'dual-tasking' studies²⁵ (VanPatten 1990 and Wong 2001a) have suggested that learners' ability to both recall propositional content and note occurrences of 'form' improved with hours exposure to instruction.

This sub principle would also find support in Williams (1999), which found, by analysing the production and learner-talk of 8 classroom learners, that the degree and type of self-induced attention to form was positively correlated to proficiency level and the nature of the activity (see section 2.2.4 for further discussion of developmental stage influencing input processing). However, this hypothesis would perhaps benefit from a sustained laboratory-based research agenda, investigating issues of implicit learning and memory.

e) *Principle 1f: The Sentence Location Principle*

*'Learners tend to process items in sentence initial position before those in final position and those in medial position'*²⁶.

²⁵ Robinson (1995) recommends the use of the dual task paradigm to research attentional demands in SLA, though it is acknowledged here that the term 'dual-task' is simplistic to describe language input processing.

²⁶ This issue is related to perceptual saliency, which is discussed in section 2.2.4 (Challenges to the theory).

Barcroft & VanPatten (1997), using immediate sentence repetition with beginner learners of Spanish found that utterance-initial position (and stress) rendered items significantly more salient for learners. Similar studies, Klein (1986) and Rosa & O'Neill (1998), are also cited by VanPatten to support this corollary.

This corollary would suggest that in French, where subjects or temporal adverbs appear in utterance initial position, learners will detect them before other items (e.g. before morphosyntactic features in sentence medial or final positions). However, this corollary is not discussed further as VanPatten (2004) specifies that it only applies "all other processing issues being equal" i.e. when communicative value is held constant. As discussed previously, this cannot be done with the target linguistic features of this study given that, according to IP, verb inflections have lower CV than subjects and temporal adverbs for communicating person, number and tense. In addition, in terms of this study, it was not possible to place verb inflections in utterance initial position during PI.

2.2.4 Challenges to the Theory

There are several challenges to IP, some raised by VanPatten and other SLA researchers and others by the author. None of these issues can be fully pursued in this study but they are mentioned here to acknowledge that the choice to use PI to explore an input-based approach to grammar pedagogy does not entail a complete acceptance of the theoretical framework claimed to support it.

a) L1 input processing

First, there is a lack of reference to literature regarding L1 processing. VanPatten (1996) defends this on the grounds of differences in conceptual and linguistic requirements between L1 and L2 learners and the different completeness of their learning. He also claims that the L1 literature does not address *why* learners process certain items and not others. Clearly the notion of 'search for meaning before all else' is highly congruent with the influential notion of top-down processing from early L1 reading researchers (Smith 1971 and Goodman 1968). However, more recent work suggests that L1 input processing is multi-resourced and highly complex (e.g. Perkins 1998 reviews studies which have emphasised the importance of bottom-up processes in reading). Here, it is acknowledged

that any model of input processing requires more detailed scrutiny of other input processing literature. One promising avenue for this is beginning to emerge: VanPatten (2004) draws parallels between IP and another *L2* input / parsing-based theory that does suggest a coalition of mechanisms are involved in input processing which, in turn, makes some positive reference to PI (Carroll 1999, 2001²⁷).

b) Modality of input processing

IP literature makes rather arbitrary references to how PI relates to both aural and reading processes. For example, VanPatten (1996) comments "...the communication goal of the learner is ...to understand what the speaker said. (Although input may be written, we will focus here on aural input)" (p17); "in any event, Leow's stimulus was written input, whereas our primary concern here is the processing of aural input" (p28) and Cadierno (1995) points out "Preterit forms were always presented first via written input and then via oral input" (p183). However, no satisfactory explanations are offered for these decisions. Similarly, the modality of the tests in most PI studies is often not systematic²⁸.

IP theory requires further refinement regarding the role of the modality of processing, and more connections could be made with related research investigating whether there are differences between auditory and visual processing systems, for example: L2 research (Wong 2001a, Bernhardt 1991) and cognitive psychology literature (Wickens 1989²⁹, Neumann & Sanders 1996, Wijers et al. 1996).

The current study uses both aural and written input in the teaching intervention, in line with other studies and the PI guidelines. The tests broadly assess whether PI has differential

²⁷ Carroll's (1999 & 2001) "Autonomous Induction Theory" posits that processing and restructuring is multi-level (morphosyntactic, acoustic-phonetic and semantic) and is facilitated by a coalition of resources (e.g. parsing strategies, negative evidence, UG, generalising mechanisms, negotiation of meaning). One argument in particular (that learning is 'a process which takes place whenever a parse fails. It is designed to restructure parsing procedures to deal with a novel input datum' (1999 p365) is compatible with PI, which is designed to induce a failed parse.

²⁸ Salaberry (1998) also suggests that the issue of modes of instruction and testing should be more carefully controlled in the studies, pointing out that in VanPatten & Cadierno 1993 the interpretation task was aural, whereas the production task was written

²⁹ Although Wickens (1989) is cited in VanPatten (1996) to support the idea of multiple pools of processing resources, VanPatten does not satisfactorily acknowledge that the dimensions of one resource pool are *auditory* versus *visual*.

impacts according to modality, though detailed exploration in terms of the impact on different language forms in different modes was not possible.

c) PI and syntactic input processing

IP assigns a vital role for SLA in learning connections between surface forms and their semantic content. This raises questions regarding the role of PI for abstract syntactic properties. Several studies have suggested that PI may have beneficial effects on aspects of syntax (position of direct object pronouns in Spanish). Furthermore, although IP is not conceived of within a UG framework, VanPatten (1996 & 1997) appeals to the generative argument that verbs must 'raise' in order to acquire inflection for person, number and tense (e.g. the English dummy verbs 'do' and 'have', or verbs before the negator in French), and suggests that by promoting intake of these meaning-bearing features, learners would also acquire the verb movement that 'carries' them³⁰. It was considered that this issue was beyond the scope of this study, as it would have required even more classroom time to test the development of negation as well as inflection.

VanPatten (2002) suggests that processing across clause boundaries (e.g. for the Spanish subjunctive) may increase the cognitive load as learners must retain information in working memory whilst continuing to process on-line³¹ thus impeding the processing of forms with low CV. Again, these issues have not yet been directly pursued by PI studies.

d) Perceptual saliency

VanPatten acknowledges that issues of perceptual saliency are of relevance to how and what is processed in the input e.g. location in the utterance (see sub-principle 1f), acoustic saliency (including prosody and boundedness), input frequency and embeddedness. Lee & VanPatten (1995), in an earlier version of IP principles, attempted to make statements about the interaction of perceptual saliency ("phonological properties (e.g. melodic contours, intonational patterns)" p97). There are numerous mentions of phonetic saliency (including

³⁰ It is acknowledged that both these syntactic features (movement of objects and verbs) 'carry' semantic meaning (e.g. number, person values) and so may lend themselves to instruction which highlights these properties. The effect of PI on truly abstract properties (i.e. features with no semantic value, such as grammatical gender) is not discussed satisfactorily in the IP and PI literature. It is suggested in VanPatten (2004) that these features may develop very late or not at all.

³¹ Such a notion finds support from other studies (e.g. Pienemann 1998 and Myles 1995).

phonetic distance) in VanPatten (1996 & 1997), though the incorporation of this into a more complete model "is left to future research for the time being" (1996 p140).

Although some potential avenues for future research are suggested, to date IP does not offer testable hypotheses regarding how perceptual saliency actually serves input processing and SLA (Sharwood Smith 1991 & 1993). Here, it is noted only that the target inflections in this study have several characteristics that, according to IP literature, may adversely affect how learners process them (see sections 1.6.3). Intensive instruction on a limited number of features holding communicative value and/or perceptual saliency constant in order to investigate the interplay of these two concepts would not have been ecologically valid in this classroom study.

e) *Reliance on output evidence: over-modularisation of IP*

The majority of the evidence advanced to support IP is output phenomena (though there is some evidence from introspective and input / interpretation studies). Reliance on output phenomena as evidence is typical of discussions of SLA (e.g. Harley 1994). It is particularly problematic in IP because in some places VanPatten suggests that output phenomena do indeed reflect processes involved in input processing (e.g. VanPatten 1996 p19, p22, p29) yet in other places he is keen to suggest his model does not account in any significant detail for e.g. the developing system, restructuring or output phenomena (e.g. VanPatten 1996 p147, VanPatten 2003 p5³²). Such modularisation has been partially accepted in this thesis in order to evaluate the usefulness of an applied theory, even though it can result in "leaking" between explanations of different modules, also acknowledged in Ellis (1995), Hatch, Shirai & Fantuzzi (1990), Bernhardt (1991 p71). As Ellis (1995) concludes, theories should not be evaluated in terms of issues they do not claim to embrace (VanPatten frequently reiterates that IP aims to focus researchers on how learners attend to input data to develop their language). These phenomenological issues cannot be developed further here.

³² In VanPatten (1997 & 2003), brief reference is made to Slobin (1979), Anderson & Shirai (1994) and Pienemann (1998) to account for output data and production constraints.

However, it will be seen in subsequent chapters that it is of relevance to note that IP could, in particular, benefit from further refinement regarding the impact of learners' 'developmental system' on their processing of input (see figure 1.4 and appendix 7). Possibly as a consequence of this, VanPatten and colleagues have maintained relative silence on *when* PI will be effective in terms of different learners' developmental stages (a difficulty facing many FonF/S approaches (Doughty & Williams 1998, Ellis 1993)). Although this study has adopted its own criteria for this matter (see section 1.6.2 and appendix 6), it is suggested that this is an important area for future research and it is discussed further in the next section.

2.2.5 Effects of developmental stage on input processing (and when PI may be effective)

The idea that characteristics of input processing change according to the developmental stage of the learner finds support in, for example, Corder (1978), Pienemann (1998), Lightbown (2001) and Robinson (1995). VanPatten also acknowledges this possibility. For example VanPatten (1990) suggests "there is also the argument that intake is structured by the learner's current grammar" (p297). VanPatten (1997) proposes that "as learners progress in acquisition, their developing system takes on a greater role in sentence processing" (p97). VanPatten (1996) claims:

"the developing system may influence input processing itself by allowing or disallowing certain form-meaning connections to be made...input processing itself is not guided solely by universal and invariant strategies; as language acquisition progresses, input processing is also guided by information contained in the developing system itself. That is, at some point, the knowledge stored in the developing system is utilized during on-line input processing..." (VanPatten 1996 p 31-39).

VanPatten (2004) suggests that the state of the learners' developing system is closely associated with the availability of resources during online processing (see section 2.2.3d regarding the 'availability of resources principle'³³).

"Just what provides for the availability of processing resources? One obvious answer is proficiency level and the nature of learners' ability to

³³ He also suggests task demands (including features of the interaction) interact with the availability of resources.

access lexical items they have already incorporated into their developing linguistic systems."

Wickens (1989), whose conceptualisation of a limited capacity processor is central to IP, suggests that parallel processing can occur when one (or more) of the tasks involved is automated. An example in terms of input processing language items of low CV, such as verb inflections, may be when the stage of the developing system is such that comprehension of lexical items or overall sentential meaning does not require all attentional resources, or when the developing system has already incorporated some representation of these items.

However, the role of the stage of the developing system remains hazy in VanPatten (2002, 2003 & 2004). For example, "one must keep in mind how effortful comprehension and processing are for beginning and even intermediate learners" (2004) (it is unclear whether the term 'beginners' refers to the number of hours of instruction or learners who are 'beginning' to process certain forms). Similarly, the guideline "teach only one thing at a time" appears to be flexible according to the level of the learner (see VanPatten 1996 pp61-2 and Benati 2001). Given the claimed pedagogical applications of the theory, this issue requires more explicit treatment in the literature (for example, regarding the likelihood that the developing system interacts with input processing in the very initial stages of SLA).

It is acknowledged that longitudinal laboratory studies are needed to explore the development of multiple processing over time to inform a "theory of the *growth of processing mechanisms*" (Pinker 1982 p667 my emphasis). Nevertheless, in this study using two classes from different schools may offer some insight, in pedagogical terms, into the influence of different developmental stages on the effectiveness of PI.

2.2.6 Summary of section 2.2: IP theory

This section discussed IP theory's central claim that for a feature in the input to be incorporated into a learner's developing system, the learner must interpret the meaning of that feature (i.e. detect it). IP suggests that this is not always likely due to certain characteristics of natural languages (e.g. redundancy) which force learners' limited capacity attentional resources to detect language features selectively on the basis of their

Communicative Value. It has been suggested that IP can offer pedagogically usable predictions, as in the following French L2 examples: with a limited capacity model of attention which selects items in the input for further processing on the grounds of their communicative value, IP would predict that early learners of French will not detect the semantic meaning available in verb inflections for person, number and tense. Principles 1a-1d would suggest that subjects (noun or pronoun), temporal adverbials and other semantic cues (e.g. lexical items to guess the meaning) will be used by learners to detect person, number, and tense rather than verb inflections. Principle 1e implies that learners will not begin to pay attention to verb inflections until other elements of the sentence are easily processed. Principle 1f suggests that, all other things being equal, learners will process the subject before the verb in SV word order, and an utterance initial adverb before other items. This section acknowledged that IP requires further refinement in several respects, particularly regarding the impact the developmental system may have on how attentional resources are allocated to process input. It was proposed that this study may contribute to the understanding of the developmental stages at which PI may be effective for the early learning of French verb inflections as it was carried out in two classes from different schools.

2.3 An evaluation of Processing Instruction in terms of Input Processing theory

Wong (2002), amongst other PI researchers, suggest "PI is theory-driven instruction because activities in PI are directly informed by a model of IP" (p17). This section argues partly in favour of this statement, particularly for the referential activities, though it indicates various aspects of PI which are less well supported by an IP framework but constitute intuitively appealing pedagogical practice. Some additional evidence, collected for the purpose of this thesis, is also mentioned, which provides further theoretical and empirical justification for PI³⁴.

³⁴ Sanz & VanPatten (1998) acknowledge that the implications for the classroom perhaps have a more substantial theoretical base than is contained in much of the actual IP and PI literature.

2.3.1 *Explicit Grammar Instruction (EGI) in terms of IP*

In line with many FonF/S approaches, PI provides some description of the target form and its function (e.g. "when you refer to 'we' in French, the verb has to end in 'ons'"). The rationale for this in IP theory is not immediately clear. IP principles do provide a rationale for giving learners information about the ineffective input processing strategies to avoid and why learners tend to use them (e.g. "try to notice the 'ons' - you probably don't notice it because the 'nous' already tells you who is being talked about"). These explanations are basically a learner-friendly expression of the relevant principle with reference to the specific language features. VanPatten (1993) suggests that such explanations raise learners' sensitivity to features in the input and that, in turn, increased comprehension of input may develop language competence³⁵. Cadierno (1995) (replicating VanPatten & Cadierno 1993) draws on two of the roles for EGI suggested by Terrell (1991): 1) as an advance organizer to aid comprehending and segmenting subsequent input which will then assist in the development of competence 2) as a meaning-form focuser that aids the learner in establishing a meaning-form relationship for morphologically complex forms. IP literature does not refer to work such as Jacoby's (1983) notions of data-driven and conceptually-driven processing (reviewed and endorsed in Robinson 1995 and Schmidt 2001).

a) *Deductive learning*

The use of the EGI *before* the practice examples shows a preference for deductive language learning, though VanPatten has not directly addressed this. It is probably driven by the notion of 'orientation' within attention as defined by Tomlin & Villa 1994³⁶. They suggest that it can help learners to be made aware before receiving input that, for example, they are about to be required to overcome a specific processing tendency. Further evidence for this can be found in studies such as Reber et al. (1980)³⁷ which showed that when relatively simple rules underlie a very complex set of stimuli, explicit presentation of the rule system

³⁵ It is acknowledged here that giving learners explicit reasons why they mis-process items suggests that some conscious understanding *is* helpful, despite claims by VanPatten (1990 & 1994) that input processing "may or may not be conscious" (my italics).

³⁶ Tomlin & Villa's (1994) definition of attentional concepts *is* adopted by VanPatten for other purposes.

³⁷ Studies by Reber are used by VanPatten for other purposes.

before examples leads to better learning than examples presented before rules, examples only or rules only.

b) The level of explanation

The issue of *complexity* of the target language feature and how this influences the style, length and detail of the EGI is not informed by IP. It seems likely that previous studies have used teachers' knowledge of the specific learners, their productive interlanguage and their familiarity with metalanguage to design the wording of the EGI. This approach has therefore also been taken in this study.

c) Role of L1

Published examples of EGIs show that there is often some comparison of the L2 with the L1. For example:

Ellos <i>observan</i> a Marcos	Marcos los <i>observa</i>
(They observe Mark)	(Mark observes them).
(VanPatten 1996 p72)	

This suggests a belief that the L1 can be a source of knowledge and possible error if learners use it to process the L2 (in some ways reminiscent of contrastive analysis).

However, this does not find theoretical support in IP. For example, the expanded model of IP in appendix 7 suggests that learners do not use the L1 to assign grammatical roles, or to attend to and detect features in the input during online processing. VanPatten's (2004) suggestion that the L1 only has a role *after* other processing strategies have been used is probably overly simplistic. It contradicts those e.g. Kellerman (1983), Andersen (1983), Carroll (2001) who argue that many aspects of learners' L1 (e.g. phonetic, prosodic and morphosyntactic characteristics) probably interact with online input processing. Further discussion of this issue is beyond the scope of this study as investigating just one language pairing (English - French) cannot inform the IP framework on the possible role of the L1.

d) Minor responsibility for overall effects of PI

Regardless of the strength of the theoretical base for the EGI in IP, several studies have claimed to have isolated and investigated the effects of the EGI component in PI. VanPatten & Oikennon (1996), Benati (2004) and others in VanPatten (2004) indicate that

actually EGI has little or no impact on the measures, finding that PI without EGI has similar effects to PI with EGI.³⁸

2.3.2 *Referential Activities in terms of IP*

Referential tasks focus learners' attention on the linguistic target in the aural and written input in such a way that they are *forced* to make form-meaning connections. For example:

According to the following sentences, underline who does the activity:

Le chat / les parents	promène le chien
Le bébé / les parents	pleurent beaucoup
L'enfant / les hommes	travaillent dans une banque

IP suggests that normally learners would focus on the lexical items and so decide, wrongly, that the first of these sentences states that parents, not the cat, walks the dog. If learners wish to succeed in this task, contextual / semantic processing is not a reliable strategy and so they must pay attention to the verb inflection and its link with number in the noun phrase. This is in line with the objective of over-coming learners' tendencies to follow Principle 1a: '*Learners process content words in the input before anything else*'.

It might be thought that referential activities have a surface resemblance to behaviourist informed approaches i.e. pattern practice with the reward of communication to reinforce the behaviour. The main differences are that in PI learners do not overtly *produce* the patterns and each item requires the processing of meaning and cannot be successfully completed by the mechanical repetition of patterns³⁹. In order to achieve this, the target feature is juxtaposed with another (often structurally or functionally similar) form, encouraging learners to recognise the defining feature of the structure i.e. what is it about feature *x* that makes it have a different function to feature *y*? For example, PI activities may help learners to unpack *j'* from the 'chunk' *j'aime*, by contrasting it with *il aime*.

³⁸ Two concerns have not yet been addressed by these studies: 1) the EGI-only control group has much less exposure to the target feature than the other groups and 2) the possible interaction of the exact EGI from the PI package (i.e. explicit instruction on interpretation strategies) with output-based instruction.

³⁹ Whether all PI activities in previous studies have always achieved this is debatable – see Wong (2001b) where learners could just listen out for a meaningless cue to interpret the causative *faire*.

However, PI researchers tend to refer to these comparison features as 'distractors' (e.g. VanPatten 1996 p99) and learning gains in them are not analysed, suggesting that it is not considered that these structures may also get reinforced. In fact a close inspection of the literature is required to reveal that juxtaposition of two contrasting forms (and their meanings) is a defining feature of referential activities, as without it learners can rely on mechanical repetition of one target form without being forced to connect it to meaning. Cadierno (1995) noted: "the presentation of the preterit tense forms involved presenting the preterit forms themselves and the contrasting those forms with present verb forms" (p183). VanPatten (2002) describes Buck (2000) as investigating: "the acquisition of present continuous (*vs.* the present progressive)" (my emphasis). VanPatten & Wong (2001) and VanPatten (2002) make clearer statements regarding this issue, possibly due to misunderstandings in previous studies, for example:

"Allen's [2000] PI activities do not force learners to distinguish between ... causative or non-causative faire. Thus, one cannot be sure of what the participants were actually learning. It is worth pointing out that in VanPatten and Cadierno, we mixed SVO and OVS/OV sentences in the materials so that learners could not apply some "mechanical" strategy of completing the activities; they had to pay attention to the sentence in order to determine which word order was being used and who did what to whom. In Cadierno, she mixed tenses up in her referential activities so that learners had to rely exclusively on the verb ending in each sentence to determine temporal reference (past, present, future)." (VanPatten 2002 p782)

Benati (2004) also emphasises this feature of referential activities:

"The [future tense] forms were contrasted with the present tense verb forms. Two things were especially emphasized:

1. the differences in acoustic stress between future and present verb forms;..."

Studies from cognitive psychology, e.g. Lee & Magill (1983), Lauer, Streby, & Battig (1976), suggest that exposing learners to items juxtaposed with contrasting items helps the learning process. It is perhaps therefore surprising that the IP literature does not draw upon such sources even though this is probably a defining feature of referential activities and may well be at the heart of the effectiveness of PI.

2.3.3 *Affective Activities in terms of IP.*

In affective activities the target feature is often put at the start of the sentence, which is in line with Principle 1f, the sentence location principle.

Other aspects of affective activities are less directly supported by IP. Unlike in referential activities, affective activities do not force learners to contrast the target feature with another and so the necessity for right or wrong answers is removed. For example, (material in square brackets is mine):

Each sentence corresponds to something that you might do to your parents. Check which ones apply to you. Compare your responses with a classmate⁴⁰

- 1) Los llamo con frecuencia por teléfono
[them call I with frequency by phone / I phone them often]
- 2) Los visito los fines de semana
[them visit I the ends of week / I visit them at the weekend]

(VanPatten & Cadierno 1993 p57)

The intention is that learners have the opportunity to 'see' the target linguistic feature (in this example, the position of object pronouns in Spanish) in some activity that refers to their own world. This is, presumably, derived from the notion that once learners have made the correct form-meaning connection in the referential activities, they are more likely to detect the meaning of the target forms when they see/hear numerous exemplars of them in the affective activities. However, the learners do not *have* to process the target feature (and in the example above, the task rubrics even tell the learners who the grammatical object of the statements is). Such a task only forces them to detect the meaning of the lexical verb and the temporal adverb - any processing of the target form would be *incidental* to the task⁴¹. Although in affective activities the form-meaning connection is sometimes explained in a metalinguistic commentary, learners may or may not read this, and may or

⁴⁰ Note the instruction "share your responses with your classmate", contradicting the claim in VanPatten & Cadierno (1993) that learners did not produce the target feature; see also Wong (2001b) (p26). This rubric was avoided in the current study.

⁴¹ A reviewer for Terrell (1991) raised a similar issue: why should learners attend if the meaning of the target form is still redundant?

may not understand it, let alone actually attend to the feature. There is no guarantee that 'detection' will occur, even though this is essential according to IP theory.

These activities probably indicate a general inclination towards embedding PI within a broadly Communicative Language Teaching curriculum (CLT) (Lee & VanPatten 1995). In fact, affective activities resemble 'enriched input' in that learners are exposed to numerous exemplars of the target feature but any processing of the target feature would usually be incidental to the task. If they are to be perceived as driven by IP theory, affective activities require some statement regarding issues of voluntary / involuntary and/or intentional / incidental learning (e.g. Eimer et al. 1996). VanPatten (2004) acknowledges that "perhaps this aspect of activities development needs to be strengthened or better yet the roles of each within PI (or any instruction, for that matter) need to be investigated".

The affective activities component was retained in this study in order to maintain treatment fidelity with previous studies. It is discussed in section 2.4 how the comparison treatment in this study was similar to affective activities (in that any processing of the target features would have been incidental to the task) and how this study therefore goes someway to exploring the role of affective activities.

2.3.4 The Guidelines in terms of IP

The following commentary summarises the arguments presented in the IP and PI literature for each guideline, and provides some additional support from other sources⁴².

i) Teach only one thing at a time

This advocates the breaking up of paradigms, in contrast to "traditional instruction" (the comparison treatment in previous PI studies). Cadierno (1995) cites Terrell (1991) to suggest that many examples of one form-meaning relationship help learners to reinforce this connection.

⁴² See section 2.1.4 for a presentation of all the guidelines together

This concept is intuitively appealing but further support could be sought from cognitive psychology where the concept of a limited capacity working memory would suggest learners could not process whole paradigms. The advice is also endorsed by Doughty & Williams (1998) and error-correction studies (e.g. Doughty & Varela 1998, Mackey & Philp 1998 and Lyster & Ranta 1997) where recasting of diverse forms did not lead to take-up).

It is pointed out (though not acknowledged by VanPatten) that this guideline can never be true for referential activities as they depend on contrasting pairs of features.

ii) Keep meaning in focus

This guideline is motivated by the over-arching statement in principle 1. It reminds materials designers to force learners to attend to the meaning of the target items. In addition, it is probably motivated by a desire to adhere to CLT's tendency to emphasise meaning over form.

iii) Learners must do something with the input

VanPatten (1996) explains this is motivated by the need to have some 'proof' that learners have attended to the meaning of the target item, rather than merely requesting or suggesting they do so.

It is noted here that this guideline is followed in referential activities, though not in affective activities (see section 2.3.3).

iv) Use both oral and written input

VanPatten (1996) comments "a combination of oral and written input is a response to claims...that some learners like to 'see' language while others don't...and is not tied directly to the principles of input processing" (p68). VanPatten's appeal to 'learner styles' is not elaborated, and, as discussed in section 2.2.4b, the issue of modality is dealt with hazily in IP and PI. It is suggested here that this guideline is mainly craft-driven and/or has perhaps stemmed from the intuitive appeal of learners' own comments.

v) *Move from sentences to connected discourse*

This guideline is motivated by IP's premise that a great deal of processing occurs at sentence level. However, principle 2 (appendix 8) does not reinforce this guideline as it suggests that event probabilities, contextual cues and lexical semantics, all of which could be available at discourse rather than sentence level, can override the 'first noun strategy'. Furthermore, the definition of 'sentence' is not clarified in the IP literature, though this is particularly necessary in terms of aural input processing. The guideline probably refers to providing one exemplar of the target feature per referential task 'item'.

vi) *Keep the psycholinguistic processing mechanisms in mind*

This is the only guideline to make explicit reference to IP theory and is described as "perhaps the most important" (VanPatten 1996 p67). It reminds materials designers to bear in mind the processing strategies learners tend to use (i.e. the IP principles) and to develop activities which encourage learners not to use them where necessary.

In summary, as VanPatten (1996) acknowledges, "only the second, fifth and sixth guidelines are related in any particular way to issues in input processing and psycholinguistics discussed [earlier]" (p67). (In addition, it was noted that some aspects of PI materials are inconsistent with guidelines 1, 3 and 4.) The commentary above implied that these guidelines probably originated for the purposes of teacher training and materials design (see Lee & VanPatten 1995, chapter 5). It is suggested here that they constitute 'intuitively appealing' rather than 'theoretically-driven' teaching advice (particularly as). This interpretation is reinforced by VanPatten's comment that "variations in the application [of the guidelines] from lesson to lesson may occur" (p67), presumably depending on the teacher's view of the learners' characteristics.

2.3.5 *Summary of 2.3: Evaluation of PI in terms of IP*

It has been argued that IP may inform decisions about which language features to focus learners' attention on, and why.

It has been suggested that some motivation for EGI can be found in IP literature, though some aspects of EGI are not clearly supported and others are even contradictory to IP.

However, several studies have shown that EGI has no significant impact on the effects of PI.

The design of the referential activities is clearly driven by the need to encourage learners to overcome input processing strategies laid out in IP.

The design of the affective activities is partly motivated by IP's 'sentence location principle' (in that the target feature is sometimes placed at the start of each utterance). However, in other respects, the current IP literature does not satisfactorily justify them. They certainly strive to keep meaning in focus and maintain learners' interest, constituting sensible practice and compatibility with CLT. It was suggested that affective activities are more akin to an input flood (a type of enriched input), and this will be discussed in the next section as it partly motivated the comparison intervention in this study.

A few PI Guidelines have links with IP theory, and others are driven by intuitively appealing 'craft knowledge'.

Although the PI package is not an exact operationalisation of IP, this is not necessary in order to carry out a classroom-based study. Parts of PI, and particularly the referential activities, represent some of the key notions in IP theory in a package that is possible to recreate in the classroom and which is intuitively appealing. PI materials are therefore considered as one possible operationalisation of IP.

2.4 The comparative dimension: Processing Instruction versus Enriched Input

This section provides a rationale for the comparison of PI with Enriched Input instruction (EI), a brief description of EI and a summary of the major similarities and differences between PI and EI. The lessons in a non-active control class are discussed in chapter 4.

2.4.1 *Previous studies cannot inform us about the effects of the structured input in PI*

The research questions in the original PI study VanPatten & Cadierno (1993) were:

- "1. Does *altering* the way in which learners process input have an effect on their developmental systems? (my italics)

2. If there is an effect, is it limited solely to processing more input or does instruction in input processing also have an effect on output?"

It is suggested here that these research questions could not be answered by that study as there was no comparison of PI with learner exposure to 'unaltered' input. Instead, PI was compared with output practice. And yet the researchers concluded:

"In terms of our research questions, we took our results to mean...that altering the way learners process input could alter their developing systems. The processing group showed evidence of this on both interpretation and production tests." VanPatten (2002) (p771).

Figure 2.1 summarises the key characteristics of the comparative dimension of most PI studies (summarised in appendix 4).

PI group	Output / traditional group
Explicit grammar explanation	Explicit grammar explanation
Warning about common errors, example and explanation of erroneous processing strategy	No warning about common errors, no example, no explanation of erroneous processing strategy
Structured listening and reading activities requiring learners to detect target feature in input	Occasional examples of correct forms to read or hear at the start of each output exercise.
Feedback re. accuracy of form-meaning connection (i.e. error correction during interpretation)	Error correction not always reported (or variable not controlled). Probably some during production tasks.
No intended speaking or writing of target structure	Carefully structured speaking and writing practice

Figure 2.1 : A comparison of PI and the output / traditional instruction in PI studies

These studies have compared one way of structuring input (PI) with various ways of practising production, in terms of their impact on learners' subsequent interpretation and production. Even Salaberry's (1997) study, which sets out to avoid "several methodological problems" of previous PI studies and "test the hypotheses in IP", compares PI with an *output* based FonF. Such studies do not cater for the possibility that IP's claims that detecting form in the input is necessary for SLA may be unfounded.

To explore more precisely IP's claims about what learners detect in the input, it is necessary to compare referential input tasks with other input tasks that do *not* intend to force learners away from the IP strategies, and to explore whether this results in marked differences in subsequent comprehension and production. Although no PI study has yet done this, VanPatten (1994) highlighted the same point:

"the critical questions that research needs to address are *if* a learner can attend to form while attending to meaning when processing input, *if so, what kind of form, and under what conditions* this is and is not possible. Thus, my claim is that research on attention in SLA must be inextricably tied to research on **comprehension**. In this kind of research, subjects must attend to input for meaning, they must know that they are supposed to attend to meaning and the task they perform must measure their comprehension in some way or another" p33. [VanPatten's italics, my bold]

VanPatten emphasises that PI is *not* another form of comprehension-based learning and that it is "critically different" to other input-based treatments e.g. "textual enhancement, recasts, input flood" (VanPatten 2002 p767). VanPatten suggests that research has not found convincing support for input enhancement (as discussed in section 1.5.3). About input floods he states: "[Bardovi-Harlig 1995 and Trahey & White 1993] are probably drawing learner attention in some way with input flooding, but the technique does not address issues of how a form or structure is processed (or not) to begin with" (VanPatten 1996 p305-6).

Such claims clearly need empirical investigation and this study compares PI with a version of the instructional approaches from which it is claimed to be distinct: an input flood with EGI.

2.4.2 The conceptualisation of incidental/ implicit versus intentional/forced learning

As laid out in section 2.2, VanPatten's IP states that detection (i.e. understanding the meaning) of the target features is essential to learn them, explaining why he suggests that PI is more beneficial than input floods or enhancement techniques. He claims that "subliminal learning" does not happen and VanPatten (2003) refers to, amongst others, (Dienes, Broadbent, & Berry 1991), where *implicit* processing (i.e. without attention) of letter strings did not allow learners to judge the grammaticality of new strings.

However, Segalowitz & Lightbown (1999), Harley (1994) and Truscott (1998), amongst others, review the controversial nature of the term 'implicit learning' - is it with or without: attentional capacity, conscious awareness, intention to learn, recollection, intention to remember and / or noticing?⁴³. As this is a classroom study, it was impossible to specify the exact nature of the learning that was happening during online processing. Indeed, Hulstijn & DeKeyser (1997) explore how such concepts can be better operationalized in the laboratory and N. Ellis (1999) notes, "some issues for example those concerning the roles of attention or consciousness in learning, can only be properly conducted in the laboratory" (p33).

In this classroom study the principal defining characteristic of Enriched Input was that the tasks forced noticing of the *lexical* items in the input (i.e. nouns, adjectives, adverbs and/or verb stems). Therefore, any detection (i.e. interpreting the meaning) of verb inflections would have been *incidental* to the task set. Evidence that interpretation or production of verb inflections improved following such tasks may indicate that simultaneous processing of lexical items and items of low CV had occurred, contradicting IP principles 1a, b, c and d. In contrast, in the PI referential activities detecting the target verb inflection was essential to the completion of the task. Learners therefore had to 'intend' to interpret the meaning of the inflection.

2.4.3 Contextual & methodological rationales

There was also a more practical, educational rationale for the design of the EI. Four characteristics of EI meant that it was similar to listening and reading activities often seen in year 9 classrooms and textbooks, as presented in section 1.4. These were: metalinguistic information, non-structured input, high frequency of a particular form and promotion of semantic processing (i.e. the verb inflection was non-essential to the completion of the task). This is similar to the justification in VanPatten & Cadierno (1993) (and others) for comparing PI with 'traditional instruction' i.e. to explore differences between PI and activities that typically form part of current teaching methods. As Ellis (1999) suggested,

⁴³ Doughty & Williams (1998) suggest that "two highly implicit techniques ... are the input flood and the use of tasks in which the learning target is essential for successful task completion." This would mean that both techniques used in this study could be broadly defined as 'implicit'.

form-focussed instruction research has a dual function - to improve pedagogy and to test theoretically-based hypotheses (p65). However, it is acknowledged that EI was not always similar to current classroom reading and listening activities in terms of presentation, content or sequence of activities.

A useful practical consequence of comparing PI with other input-based activities was that this format was possibly easier in terms of teacher planning, than an input versus output comparison. Thus teacher A commented that as all learners would do listening and reading activities in their experimental groups, the whole class non-treatment activities could be more focussed on speaking and writing.

An additional advantage of comparing PI with another input-based instruction is that sub-vocal rehearsal⁴⁴ (which has not been 'controlled' in previous PI studies and therefore renders the distinction between input and output instruction hazy) is more likely to be at least 'uncontrolled' in *both* groups in this study. Though this does not resolve the issue (which would have to be addressed in a laboratory using sophisticated recording techniques), it does improve the study's internal validity. In addition to sub-vocal repetition, previous studies have also not controlled for the possibility that learners' output could provide input for other learners and even auto-input⁴⁵. This issue is also more satisfactorily addressed in the current study as neither group were expected to produce the target forms overtly.

2.4.4 A comparison of each aspect of the packages

The distinction between the two teaching techniques investigated here is thus relatively small compared to other PI studies. This section outlines the differences between each intervention package.

⁴⁴ Sub-vocal rehearsal was suppressed in Papagno, Valentine, & Baddeley (1991) which found that it did indeed contribute to retention in short term memory. Bernhardt (1991) also reviews L2 reading studies to conclude "sub-vocalizing probably occurs" (p77).

⁴⁵ It is acknowledged however that in previous PI studies, even if output *did* unintentionally serve as input, the 'output' learners still did not perform as well as PI learners in interpretation tasks, and only the same in production tasks. This may be because such output did not constitute 'structured' input for the learners e.g. it contained cues of higher CV.

a) The EGI in the PI and EI groups

Both groups received identical EGI. This was to eliminate the possibility that explicit instruction could explain any subsequent differences found between the PI and EI learners' post test scores. VanPatten & Oikkenon (1996), Benati (2004) and Sanz & Morgan-Short (2004), amongst others, already provide some empirical support for eliminating this explanation. However, the control group in these studies had EGI only (it was not combined with Traditional Instruction). They therefore cannot serve to eliminate the explanation that, in PI studies such as VanPatten & Cadierno (1993), the comparison (Traditional Instruction) learners did not interpret the target items as well as the PI group because they did not receive EGI about *interpretation* errors and strategies⁴⁶. The relevant point for this study is that both groups should be given the same EGI in order to study the interaction of EGI with both packages.

b) PI referential activities and the EI comparison activities

The number of target features in the input was held constant in both types of activity but the EI activities had the following differences to the referential tasks: they did not make the verb inflections essential to the task (learners could process the input 'semantically' by focussing only on nouns, adjectives, verb stems or overall sentential meaning, though learners may have 'self-directed' their attention to the target forms); the target feature could be presented at the sentence or discourse level; the verbs were not necessarily placed in sentence initial position; items of higher CV (e.g. temporal adverbs, subjects) were retained alongside almost every target verb inflection.

c) Affective activities and their comparison activities

Despite the criticisms in this chapter regarding the lack of principled connection between affective activities and IP theory, the affective activities were maintained for parity with other studies (i.e. to create a conceptual replication of PI according to the criteria laid out in Polio & Gass (1997). However, to reduce the influence of affective activities in the comparison of PI and EI, they were similar for both groups in that any processing of the target forms would have been incidental to the task. Both types contained an input flood of the same number of target forms. In addition, the nature of the tasks was usually similar (in

⁴⁶ They usually performed equally well in production.

terms of content and aims e.g. expressing opinions). There were just three differences between the PI and EI versions of affective activities, summarised in figure 2.2.

<i>PI Affective activities</i>	<i>Equivalent EI activities</i>
Target feature's form and meaning was pointed out to the learners by means of a 'speech bubble' on the task sheet. The teacher also read this out.	No reiteration of explicit grammar instruction
Target feature placed at or near the start of each sentence whenever possible.	No structured presentation of the input
Input always at the sentence level – learners had to do something at the end of each sentence.	Input at either sentence or discourse level - learners had to do task after each sentence or after paragraph (reading) / speech turn (listening).

Figure 2.2 Comparison of PI affective activities with equivalent EI activities

Notwithstanding the differences in the affective activities outlined above, the design of the intervention activities in this study could be summarised as:

EGI + referential + affective *compared to* EGI + input flood + affective

or as:

EGI + structured input activities *compared to* EGI + input flood

2.5 Summary of chapter 2 & research questions

PI can be seen to be driven to some extent by IP theory, especially the referential activities. PI is also a coherent, replicable and intuitively appealing pedagogical package. However, PI studies to date, due to the output-based nature of their comparison groups, have not yet satisfactorily investigated the effects of the *specific* type of input-based instruction that is PI, nor, therefore, IP theory. A refined exploration of IP is not within the scope of this study either, as this would require a programme of research accessing internal and online processing during PI. Nevertheless, IP theory has suggested one way in which to investigate the use of input in the classroom for the learning of communicatively redundant verb inflections: IP theory has provided a principled framework for exploring the differential effects of altering what learners are asked to do with input, in this case

intentional (i.e. 'forced') versus incidental input processing i.e. structured input versus EI. This study's unique comparative dimension was whether or not learners were encouraged, by means of PI's referential activities, to overcome some of the processing strategies suggested in IP's principles. This study may also offer some insight into the impact of different developmental stages on the effectiveness of PI as the study was carried out in two different schools. Additional design issues, problematic in previous PI studies, are discussed in the next chapter (e.g. the measurement criteria and the length of the intervention).

Research Questions

- 1) Do structured input activities have any favourable effects compared to Enriched Input on year 9 learners' ability to *interpret* in reading and / or aurally French present and perfect tense verb inflections, as measured by a battery of achievement tests?
- 2) Do structured input activities have any favourable effects on year 9 learners' ability to *produce* in writing and / or orally French present and perfect tense verb inflections, as measured by a battery of achievement tests?
- 3) Are the same results for questions 1 & 2 maintained in *delayed* post tests, taken between 4 and 6 months after the start of the intervention?
- 4) Are the same results for questions 1, 2 & 3 obtained with learners from two 'similar' classes from different schools?

Chapter 3 Methodological Discussion and the Current Study

The research questions presented at the end of chapter 2 require the identification of causal relationships between teaching and learning. *How* to study causal links between teaching and learning using experimental approaches is discussed at the start of this chapter. Following sections lay out the experimental design used, the design and implementation of the teaching materials and the tests, scoring procedures used and an explanation of the statistical analysis procedures used. The chapter has the following structure:

- 3.1 A critical review of literature regarding the role of an 'experimental approach' in educational research
- 3.2 The current study: The participants, their settings and the build up to the experimental intervention
- 3.3 The quasi-experimental design: The intervention period
- 3.4 The language sample for the teaching materials and tests
- 3.5 Designing the tests
- 3.6 The tests
- 3.7 Monitoring the process
- 3.8 Procedures for the analysis of the achievement tests
- 3.9 Summary & original features of the study

3.1 A critical review of literature regarding the role of an experimental approach in educational research

This section describes the characteristics of the positivist tradition¹ and outlines the problems of applying the positivist paradigm to educational research. The strengths and weaknesses of an alternative paradigm ('naturalistic', broadly referring to approaches

¹ The term 'positivism' was first used by Auguste Comte, a C19th French philosopher, to refer to a philosophical position where it was believed that the methods of the physical sciences (i.e. experimental) could and should be applied to social sciences.

where there is no intentional, controlled intervention treatment) are then discussed before concluding that adapting an experimental approach can serve this study.

3.1.1 Characteristics of the positivist tradition

The adoption of an experimental approach usually indicates that the researcher tends towards rational, realist, objective, quantitative and deterministic perspectives on the nature of social science and knowledge (Burrell & Morgan 1979), and this gives rise to certain methodological assumptions (Hitchcock & Hughes 1995)². This section reviews key features of an experimental approach and explains how the influences of context in experiments need to be empirically definable and / or controllable in order to maintain validity.

a) Characteristics of True Experimental Designs

There is considerable consensus in the literature regarding the following defining characteristics of true experiments (Cohen, Manion & Morrison 2000, Isaac & Michael 1995, Hatch & Lazaraton 1991, and Goodwin & Goodwin 1996).

- i) A true experiment investigates testable and falsifiable hypotheses (Kerlinger 1970). The null hypothesis is normally that the experimental treatment (i.e. the independent variable/s: in this study, the instructional type and the class) will have no significant impact on the outcome under investigation (i.e. the dependent variable/s: in this study, the language interpretation and production measures). The alternative hypothesis usually states that the experimental treatment *will* bring about a significant difference.
- ii) To eliminate rival explanations and to maximise the variance caused by the experimental manipulation, there should be rigorous control of variables either by direct manipulation or through randomisation (*ceteris paribus*) (Fitz-Gibbon 1996). In language learning research, for example, many laboratory-based studies have used

² See Cohen, Manion & Morrison (2000) (chapter 1) and Elliott (2001) for reviews of the philosophical, epistemological and methodological debates that have characterised and influenced debates regarding experiments in educational research.

artificial languages to control extraneous variables such as prior exposure, L1 transference, influence of cognates e.g. Reber (1967).

iii) A true experimental design uses a *control* as a baseline for comparison with the group receiving treatment as this is believed to reduce researcher bias, maximise the reliability of findings and control / define the potential confounding influence of extraneous factors.

b) *Validity*

One of the principal advantages of the positivist approach can be the high internal and external *validity* of the conclusions.

Internal validity is concerned with showing that it was the experimental manipulation that was responsible for any perceived difference between groups. This requires: i) the ascertaining of homogeneity of the groups being compared (by randomisation or creating groups on the basis of a pretest, and / or taking initial differences into account in subsequent analyses) ii) using the 'same' pre and post test and iii) following a linear protocol to reduce the risk of adding uncontrolled variables. Fitz-Gibbon (1996) suggests that "experimental research... forces the researcher to operationalize the intervention - to spell out exactly what has to be done to implement the 'treatment'" (p105-6).

External validity concerns the 'generalisability' of the findings (Fitz-Gibbon 1996 p113). This means that the intervention and how change was assessed must be sufficiently documented to make it replicable. In addition, probability is applied to determine the likelihood that something true has been discovered, or that the prediction will come true next time.

The 1960s and 70s saw many positivist studies in most educational contexts, including the UK. Freedman (1978) illustrates the dominant experimental paradigm of the time in her study comparing the effectiveness of grammatical drills, rules and language laboratory materials in UK secondary schools.

Campbell & Stanley (1963) did much work defining the characteristics of experimental design. Their analysis of 16 types of research design, to evaluate how effectively each overcomes the threats to validity, concluded that the one shot case study (one group post-test-only design) was the least successful at meeting this criterion and the true experimental design was the most effective.

3.1.2 Problems of applying a positivist approach to educational research and alternative naturalistic approaches

Experimental approaches in some countries e.g. USA, Netherlands and Korea, have never been completely out of favour (Oakley 1998, Lagemann 2000). However, in UK educational research, the late 1970s, 80s and the first half of the 90s saw a move away from experiments³. This section details some of the principal technical reasons why many educational researchers (and other social scientists) have adopted naturalistic/non-interventionist approaches since the 1980s (Hillage & et al. 1998) and briefly discusses their advantages. Reference is made to language learning studies, where possible, to maintain relevance with methods for the current study.

a) Post modernist thinking

Post-modern thinking emphasised the complexity, subjectivity and context dependency of human phenomena (Schensul, Schensul & LeCompte 1999, Isaac & Michael 1995). Proponents of naturalistic research are “united by their common rejection of the belief that human behaviour is governed by general, universal laws” e.g. simple causal relationships between specific interventions and subsequent behaviours (Pring 2000 p19).

Hammersley (1987) argued in favour of gathering rich data from individual contexts using ethnographic techniques and qualitative data to describe subjective experiences and individuals’ perceptions of reality. Such approaches can explore a previously unknown phenomenon in a way that experimental hypothesis testing can not. For example, Borg (1998 & 1999) documents the complex influences acting upon language teachers’

decisions regarding grammar pedagogy and concludes that naturalistic studies give greater insight into how to inform teachers' practice than trying to identify and apply a 'best method'.

b) Political partisanship

There was (and still is, e.g. Atkinson 2000, Hammersley 1999) a feeling that the need for experimental evidence of 'what works' in education is a manifestation of political partisanship, as it could involve providing evidence post-hoc for policies or making value-judgements (for example, assumptions required for this study are that teaching MFLs is a desirable thing in the first place and that accurate target-language production should be part of the curriculum).

c) Difficulty of controlling extraneous variables & artificiality

Glass (1979) noted, after a meta-analysis of social science research, that only a third of variability in the data can usually be explained by the variables under study and that 33% of this could even be predicted given the type of research instrumentation used. His conclusion was that such results can not make useful practical impacts. The very appeal of the experimental approach seems also to be its downfall - the tight focus on a single issue can ignore the influence of other factors (Bryman & Cramer 1995). For example, Burstall et al. (1974), Freedman (1978) and Fitz-Gibbon & Reay (1982) suggested that their results may have been influenced by relationships between achievement and the students' and /or teachers' attitudes to the experimental intervention.

The classic "experimenter's dilemma" (Jung 1971) is that there is a trade-off between maintaining control over the protocol (i.e. maintaining internal validity) and reducing the artificiality of an experiment (i.e. maintaining external validity). One possible consequence of artificiality is that subjects' behaviour changes as a result of realising they are being studied, rather than as a result of the variables being manipulated (known as the 'Hawthorne Effect', (Cohen, Manion & Morrison 2000 p127). In addition, if participants find the conditions artificial or unsatisfactory they can subvert the protocol

³ A few educational researchers retained a more positivist and quantitative perspective, mainly represented

(for example Spada & Lightbown 1993 found that the control teacher had given corrective feedback and explicit rules instead of just focusing on meaning). Similarly, Harley (1989), Lightbown (1991) and Day & Shapson (1991) found that their FonF/S experimental protocol was subverted in some way, illustrating that comparison treatment/s are difficult to 'control'.

It is often found (Isaac & Michael 1995, Goodwin & Goodwin 1996) that *post facto* correlational designs are anyhow the only option in circumstances where the 'more powerful' experimental method is not possible: when randomisation is not possible; where *ceteris paribus* is highly unrealistic and artificial (e.g. teachers do not teach the same pupils for extended periods of time (Mitchell 2000a and LeCompte & Schensul, 1999 p75); and/or where only small data sets can be obtained.

d) Lack of appeal to stakeholders

A lack of apparent usefulness to practitioners and policy-makers of educational experimental research findings may have contributed to their decline⁴. The research may have been unable to offer firm conclusions because of insufficient planning regarding the control group e.g. Harley (1989) (where the control group was exposed to some of the experimental treatment) or Aljaafreh & Lantolf (1994) (where the control group had less hours tuition than the experimental group). Findings have sometimes been counter-intuitive, or interpreted as such e.g. Burstall et al. (1974)⁵ and Tymms (1999) (which suggested homework has no measurable benefits). Alternatively, the research revealed something that the majority of practitioners felt they carried out every day. Stenhouse (1975) and Schön (1983) favoured teacher-as-researcher, non-experimental approaches, as these could directly inform practice and develop educational theory accessible to practitioners.

by Goldstein.

⁴ Some of these criticisms can also be levelled at naturalistic approaches but they are presented here as problems particularly facing experimental designs due to the demands laid out in 3.1.1

⁵ The results showed that learning French at primary school made no statistically significant gains for learners at secondary level. However, the probable reason was lack of co-ordination between the two levels.

e) Lack of agreement regarding constructs

The generalisability of claims is reduced if constructs are not defined satisfactorily (White 1999). For example, the lack of an agreement regarding the evidence of learning is a common problem in SLA studies: Salaberry (1997) critiques VanPatten & Cadierno (1993) for their lack of narrative elicitation tasks; Schwartz (1993) refers to the lack of positive results for FonF treatment in delayed post tests; Sanz (1997) summarises the variability of measures and illustrates how variables that are frequently left uncontrolled (e.g. mode or time on task) can affect results; Spada (1997) and Norris & Ortega (2000) point out that the length of time between the intervention, post-tests and delayed post tests in FonF/S research is inconsistent; Truscott (1998) eliminates from his critique of FonF studies those with no delayed post tests⁶, those that only measured metalinguistic knowledge and those with unsatisfactory internal and external validity e.g. Spada & Lightbown (1993).

f) Ethical considerations

The ethical considerations of doing experiments in educational settings are discussed in Gall, Borg & Gall (1996), Pring (2000), and Kember (2000). Depriving learners of the perceived benefits of a certain type of instruction and teachers of the relevant training can be seen as unethical and this can subvert the experimental design. The linear protocol also raises ethical concerns if the teacher feels that the treatment is ineffective e.g. Harley (1993) found that teachers subverted the protocol as they considered the material irrelevant to the syllabus). Kember (2000) argues that as teaching is iterative, the research approach should be flexible, allowing revision of techniques during the investigation.

As a consequence of some or all of the factors above, educational researchers have tended to work with the small, non-randomised samples available and acknowledge the context-dependency of the results.

⁶ Though, as Harley (1994) points out, the insistence of evidence of progression in delayed post tests rests on the assumptions that it doesn't matter that the instruction stops and that 'internalised linguistic knowledge' is never forgotten.

3.1.3 *The disadvantages of naturalistic studies and recent educational debate*

This section lays out some of reasons why ongoing academic and political debates about research quality in the UK have continued to suggest that educational research should provide generalisable 'effective practice' advice to practitioners (Hammersley 1999, Hillage & et al. 1998, Rudduck & McIntyre 1998 and Blunkett 2000).

a) Subjective interpretation

One complaint has been that data from naturalistic studies can mean that the perception and interpretation of patterns may involve subjective processes (Gall, Borg & Gall 1996). This issue relates to the falsifiability (Popper 1968) of some claims made in naturalistic studies where 'alternative' explanations are sometimes claimed to abound. Fitz-Gibbon (1996) compares the reliability of some qualitative evidence (e.g. selective quotation) to statistical data that contradicts the implications from the qualitative analysis (p21). Page (1999) illustrates the difficulties of finding relationships in data from a descriptive study. Mitchell (2000b) suggested that teacher process data (collected during a project tracking linguistic development (Mitchell & Dickson 1997)) was gathered under uncontrolled conditions and consequently detailed causal relationships between teaching and learning were difficult to infer.

b) Lack of impact on policy/practice

Glass (1979) and Fitz-Gibbon (1996) suggest that the non-interventionist nature of educational research meant that researchers tended to debate the status-quo rather than influence decisions. Fitz-Gibbon (2000) suggests that experimentation can address the validity issues which often lead politicians to ignore research findings. Indeed, several calls have been made for increased use of experimental and quantitative designs which can provide measurable evidence to inform (or justify) policy decisions (Hargreaves 1997, Tooley & Darby 1998).

c) Paucity of studies to contribute to meta-analyses

One consequence of the decline in controlled trials in UK education is that there is not a sufficiently large body of small scale studies which could contribute to 'meta-analyses'. This is a method of combining and accumulating small studies to estimate the magnitude

of experimental effects (Fitz-Gibbon 1999). The EPPI centre's programme of systematic reviews of educational research has experienced a shortage of high quality studies with comparable research questions and with quantifiable or clear findings (Budge 2002)⁷. Although meta-analysis has been used in SLA 'type of instruction' literature e.g. Norris & Ortega (2000), Collentine (2004), one of the main messages of the Norris & Ortega review of FonF/S studies was the need for agreement regarding the minimum research design requirements for studies to make a useful contribution to such meta-analyses.

d) Calls from researchers & practitioners

The calls from policy-makers may simply be formalising long-standing calls from practitioners and applied linguists that research should offer teachers practical advice e.g. Borg (1996), Kramsch (1995), Ellis (1995) and Trim (1988). There is considerable consensus that the classroom context should be investigated experimentally despite its complexity e.g. Ellis (1999) comments "typically grammar lessons are not constructed around a single macro-option but rather involve a combination of options...From a research perspective, however, it is useful to try to tease out the relative effectiveness of instruction based on these different macro-options" (p64).

3.1.4 The adoption and adaptation of experimental approaches

The discussion so far has presented experimental and naturalist paradigms as though they were mutually exclusive, a characteristic not uncommon in the literature (particularly amongst action research methodologies (e.g. McNiff 1993, Elliott 1991). However, a number of researchers have suggested that the two perspectives be combined (Cohen, Manion & Morrison 2000, Hillage et. al. 1998, LeCompte & Schensul 1999, Ellis 1999). This final section presents arguments, using a range of general educational research and SLA / FonF/S literature, which suggest that we are now in a position to adopt relevant parts of both approaches.

⁷ The Evidence for Policy and Practice Information and Co-ordinating Centre is part of the Social Science Research Unit, Institute of Education, University of London. The current reviews in MFL teaching and learning are described at
http://eppi.ioe.ac.uk/EPPIWeb/home.aspx?page=/reel/review_groups/MFL/home.htm

a) Need for descriptive approaches

Goodwin & Goodwin (1996) suggest that descriptive and correlational designs can act as a bridge to "experimental designs yet to be considered" (p42). Findings from descriptive studies are required to inform the design and focus of experimental investigations, for example, studies exploring the teaching and learning of grammar (Mitchell & Martin 1997, Myles 2003), and studies exploring MFL, and general, effective teaching and learning literature (Naiman et al. 1978, Cooper & McIntyre 1996).

Not only are process-oriented studies needed *prior* to experiments but they are also vital *during* the experiment for triangulation purposes, to probe 'odd' findings and to identify the extent of uncertainty around the results, as identified by LeCompte & Schensul (1999): "ethnographic research directed to careful description of the program context and process is a necessary complement to quantitative research designs" (p82). The interpretation of the Burstall report illustrates the risks involved in focussing on the quantitative results of an experimental study without looking at the problem from a more qualitative perspective⁴.

b) Impact of experimental approaches

Fitz-Gibbon (1996) cites a number of controlled field studies that have directly informed policy decisions. Reynolds & Mujis (1999) and Harris (1998) are examples of surveys of a mix of process studies, post-facto correlational, quasi-experimental and experimental studies in mathematics education. The National Literacy Strategy (DfEE 1998) and the National Numeracy Strategy (DfEE 1999) are now supported to some extent by findings from experimentally oriented effective practice research⁸. Ellis (1999) argues that "as illustrated in Chaudron 1988, Doughty & Williams 1998 and Lightbown, Spada & White 1993, classroom research that is tight enough to inform second language instruction is possible" p33. Outcomes of the FonF/S experimental tradition to date are that the research agenda is being refined and some broad implications for the classroom are already being drawn (Cook 2001, Spada 1997, Norris & Ortega 2000, Mitchell 2000a).

⁸Though these were presented *after* implementation e.g. Beard & Willcocks (2001).

c) 'Micro-level' factors can explain achievement

Fitz-Gibbon (2000) suggests that less than 9% of the variability found in the results of school cognitive tests can be explained by socio-economic factors. Fitz-Gibbon (1996) also suggests that investigations on the department level, not the school level actually reveal the greatest variability in effectiveness (p32). Creemers (1994) cites evidence that suggests that factors at school and classroom level can explain 9-27% of attainment and about 20% of progress. Such arguments point to the potential worth of investigating relatively detailed aspects of instructional techniques.

d) Response to ethical concerns

It is acknowledged that if it is *known* that a certain treatment is beneficial then it is indeed unethical to deprive a control group of such treatment. However, if the effects of a specific intervention are unknown, then withholding it should not be problematic Fitz-Gibbon (1996). (The PI and EI experiments in the international literature have not, as yet, produced clear-cut findings).

In addition, treatment (in this case grammar pedagogy) is already unevenly distributed amongst pupils (see chapter 1). If experimental designs are refused on the grounds that differential treatment is unethical, then this creates the illogical position that differential teaching styles are acceptable if part of the status quo but unethical if introduced systematically as part of an experiment.

e) Overcoming some of the methodological difficulties

Some of the difficulties of carrying out experimental designs in educational contexts can be compensated for by adopting particular techniques. Isaac & Michael (1995) suggest that the 'quasi-experiment'⁹ is as near to the true experimental design as one would hope to get working with intact groups in a natural environment. However, the lack of randomisation in quasi-experimental work does not have to entail a loss of internal validity as careful pre-testing and statistical techniques are often considered sufficiently

⁹ also called the controlled field study

rigorous to compensate for the absence of entirely randomised testing (for example, Harley 1989). Potential test effects, due to having the same pre and post test, can be reduced by withholding answers, using two or more versions of the 'same' test, ensuring the time period is considered long enough for learners to forget the items and ensuring the learners had no instrumental motivation to obtain / remember the answers. Fitz-Gibbon (2000) suggests that using indicators that are both 'internally and externally valid' can help to overcome some of the difficulties of teachers having to respond to external constraints (though this was not possible in this study). The 'motivation / achievement cycle' is an important factor when investigating the effects of an intervention, and attempts can be made to control this by introducing similarly 'novel' elements in both the experimental and comparison groups. Truscott (1998) and Norris & Ortega (2000), reviews of FonF/S studies, offer some methodological advice for carrying out classroom experiments: the definition of 'form' must be clear; some kind of (semi-) spontaneous oral use should be measured (or consensus regarding achievement measures is required); delayed post tests should be administered; there should be closely monitored comparison / control groups and an effect size should be calculated. Advances in statistical methods and software allow the influence of more than one variable to be traced with much greater ease and speed (Isaac & Michael 1995, Campbell & Stanley 1963).

f) Responding to some post-modernist arguments

There is broad consensus that routes of SLA *do* appear to be fairly universal (see appendix 6). Thus it should be of interest to see if it is possible to speed up the process by isolating and manipulating aspects of the teaching. The FonF/S paradigm, in line with positivist tendencies, favours the notion that conceptual replication studies can inform the search for universal principles e.g. to see if results hold for different populations / settings / modalities (Polio & Gass 1997)¹⁰. In addition, the study of morphosyntactic development *can* be done quantitatively and can therefore feature as a measurable construct.

¹⁰ Although replication studies are not always perceived as such (see the series of exchanges between Sanz & VanPatten (1998) and Salaberry (1998), or VanPatten & Wong (2001) and Allen (2000)), this very dialogue focuses researchers' attention on methodological precision.

3.1.5 Summary of 3.1: the arguments surrounding methodological choices in the current study

In agreement with some of the literature examined here, it is argued that educational research should not concentrate solely on reporting subjectivity and describing processes. It can also seek to evaluate the effectiveness of different practices so long as some attempts are made to enhance external validity by, for example, using intact classes with their normal teachers. The fact that certain variables (e.g. socio-economic status) may have greater explanatory power regarding learner progress than those under examination here does not justify a refusal to explore the effects of other variables (e.g. particular teaching approaches). This review concluded that successful educational experimentation should follow on from and be accompanied by in-depth process monitoring. The following sections outline how the characteristics considered in this review as desirable for a successful educational quasi-experiment have been operationalised in this study.

3.2 The current study: The participants, their settings and the build up to the experimental intervention

3.2.1 The schools

The study was carried out in two state secondary schools in an English city. Both schools have above the national average A*-C pass rates. School 1 is an 11-16 mixed comprehensive in the city's suburbs, and a Language College (i.e. a school with special status and extra funding awarded on the basis of a range of language teaching and learning initiatives and successes). School 2 is an 11-18 girls' comprehensive in the city. Consent for the study was given by the schools' senior management¹¹.

3.2.2 The teachers

At school 1

Teacher A has 30 years of teaching experience at school 1.

¹¹ As the research was integrated in the regular education the school was offering, consent from individual learners was not sought.

Teacher C taught the 'parallel' set to teacher A's group, which was used as the 'non-active control' (with no intentional experimental intervention). She was in her second year of teaching and left the school at Easter, just after the post tests in this study.

Teacher D was employed to replace teacher C on a temporary contract, and had about 15 years experience as an MFL teacher. Teacher C (then D) taught the non-active control group for 3 out of 4 lessons per fortnight. The fourth lesson was taught by teacher E throughout the year. She was in her 3rd year of teaching.

At school 2

Teacher B is a French native and had taught at school 2 for about 25 years. She taught the class used in this study throughout the year and also in the previous year.

3.2.3 The learners

Three 'top set' (as assessed by the MFLs departments) year 9 classes were involved in the study¹².

All pupils had four French lessons per fortnight (though the distribution of these within the fortnight varied). Pupils in school 1 had an official homework allocation of 30-45 minutes every Monday and in school 2 of 35 minutes every Tuesday and once a fortnight on Fridays.

All the pupils were part of the first national cohort to have experienced the National Literacy Strategy (started in 1998-1999 in Primary schools). It can be broadly assumed that 'some' metalinguistic concepts were introduced to the learners in their final year of Primary school. However, the learners' metalinguistic knowledge was not ascertained at the start of the study.

¹² Year 9 learners are 13/14 year olds with 3 years (approximately 180 hours) of exposure to classroom instruction in French. 'Top ability sets' are teaching groups created by the schools – learners in them being deemed more advanced in French than, in both cases here, one other class of their peers. The decision to use these learners, and particularly those more likely to be at a higher developmental stage, was partly motivated by findings reviewed in appendix 6 suggesting that verb morphology appears to be emerging at this stage amongst some learners, and partly by pragmatic considerations (e.g. year 9 is not part of the GCSE year where external pressures may have prevented teacher participation). The decision to use 'top sets' was also taken by the participating schools.

A brief questionnaire (appendix 9) was completed by all pupils prior to the intervention, giving information about prior and extra-curricular language learning / exposure. Several learners in classes A and C in school 1 reported that they had a French penfriend with whom they exchanged letters a few times a year. Several pupils in all classes reported that they had been on holidays to France, though no pupil considered that they had communicated a great deal in French. It was considered that no pupil had significant experience of the target language outside the classroom.

3.2.4 Information about the learners prior to the onset of this study

All except 4 learners in teacher A's class had been in a 'top set' year 8 class in which I had documented some aspects of the teaching and learning of the present and perfect tenses for an action research project¹³.

All except 5 pupils in teacher C's class were previously in a top set year 8 class that had received 'the same' teaching (scheme of work, materials and teacher) as documented in the action research project mentioned above (I also observed this class once, see appendix 1c). This class was used as a non-active control (although it is unlikely there is such thing as a 'true control' in educational research, it was important to document the learning where there was no intentional experimental intervention, particularly as classes A and C were 'parallel sets' and had experienced similar FonF/S instruction in the present and perfect tenses the previous year).

I had no contact with teacher B's class prior to the start of the study in September 2001. She taught the same intact class throughout years 8 and 9. A semi-structured interview with her prior to the start of classes mainly focussed on her beliefs about language teaching, in particular grammar teaching and also details of what the learners had done in year 8, examples of some written work (including accounts of holidays in the past) and techniques she had used with them.

¹³ This project is not referenced to maintain anonymity.

3.2.5 *Preparing the teachers, the learners and myself – September to Christmas*

Teachers A, B and C were given a 2-side outline of the study, emphasising the importance of carrying on with normal teaching whenever possible and describing the expected involvement of the teachers and pupils, the broad aims of the study (to compare two ways of teaching grammar) and what the teachers / school may hope to gain from being involved. A timeline for the whole study was also given to the teachers, with details about the first third of the study from September to Christmas. Before the Christmas holiday, an individual calendar of lessons was drawn up for teachers A and B, taking into account school events that were likely to affect lessons during the intervention period. This allowed some potential problems, such as planned teacher absence, to be identified.

When I asked teacher C to participate I explained the following: the precise focus of the study and intervention materials would be withheld from her until after the study; 'normal' lessons would proceed as far as possible, which would be observed and recorded fairly frequently throughout the year; her pupils would undertake the pre, post and delayed post tests¹⁴. To my knowledge, teacher C did not make efforts to discover more details about the project (teacher A also believed this to be the case). There is little further reference to class C in this chapter: chapter 4 includes descriptions of their lessons throughout the study and analysis of their results is in chapters 5 and 6.

All three teachers preferred that I introduced myself on my first visit to classrooms. I addressed the pupils along the following lines:

'I am a French and Spanish teacher and I used to teach in [an English city]. I will be here working with you for a few months on some research being carried out at the University of Southampton. I am also doing some work with another school. I'll be here about every other lesson, at least once a fortnight and will be coming round to help and see what you are doing. I am not here to assess individuals, I am interested in how pupils your age are learning French. I'm sure you'll get used to me. I'm very pleased and grateful to be able to come and work with you and I look

¹⁴ Clearly, my presence, the recording equipment, the three administrations of the battery of tests may have affected the teaching and learning, and this is discussed during the analysis of the results.

forward to it. Do you have any questions at this stage? or you can ask me later'.

3.2.6 Collecting data prior to the intervention

A description of the study prior to the experimental intervention is given here, structured according to the two aims at this time: familiarisation and collecting contextual data from the different teaching and learning environments in each class. Appendix 10 provides details of the lessons observed and records collected throughout the study.

a) Familiarisation

Familiarisation (perhaps more normally associated with ethnographic techniques) aimed to improve certain aspects of experimental validity, for example by reducing the impact of artificiality if pupils, teachers and I could become accustomed to each other and get used to practicalities such as moving between rooms, being video and audio-recorded and classes being split into groups¹⁵). Every learner wore a lapel microphone during French lessons for about 30 minutes at least twice (in preparation for being recorded during the oral elicitation tasks). Lessons were also video or audio recorded, using a mounted camcorder in a corner at the back of the room or a digital IC recorder with a lapel microphone on the teacher. To help the learners and myself become accustomed to me being their teacher in a split class situation (see later), in October and November we split each class in half on 4 separate occasions for at least half a lesson. Every learner experienced my teaching twice. I followed the class teacher's lesson plan and made attempts to adopt some key characteristics of the pupils' normal class teacher e.g. error-correction, target language use, pace. In both schools we told the learners that they would be doing the same activities as the other group and that working in these smaller groups would help them get used to this arrangement in preparation for the next term.

b) Collecting contextual data

The second aim of this period of observation was to collect data about the instruction the learners had prior to the experimental treatment (see appendix 10 for details).

¹⁵ An additional measure taken against Hawthorne effects was that both experimental groups experienced 'novelty' to a similar extent in the design of the PI and EI materials, see sections 3.7.2 and 4.3.2.

Observation notes (see appendix 11 for an example) were made in reference to: general classroom management (e.g. time on task, homework, apparent commitment of the pupils), teacher and pupil use of target language, materials and activity types, error correction strategies and when and how the target features occurred in the lessons. Occasional notes were made regarding how some activities compared to PI and EI. Systematic observation schedules were not used as it was unknown at the time which specific issues may affect the interpretation of the test results. With both teachers it was agreed that, if necessary, I would help during the lessons (e.g. by circulating to correct work and respond to questions). On the few occasions when asked to do so, I tried to respond to pupils in a similar way to their class teacher.

I observed teacher C's lessons from late September until Christmas four times, making field notes each time and three audio-recordings. I used the same techniques as described above for familiarising the pupils with being individually recorded.

(Lessons were observed in each class in the same way between the post and delayed post tests, to help with the interpretation of the results).

3.3 The quasi-experimental design during the intervention

The following diagram illustrates the design of the study during the intervention period, late January to early March.

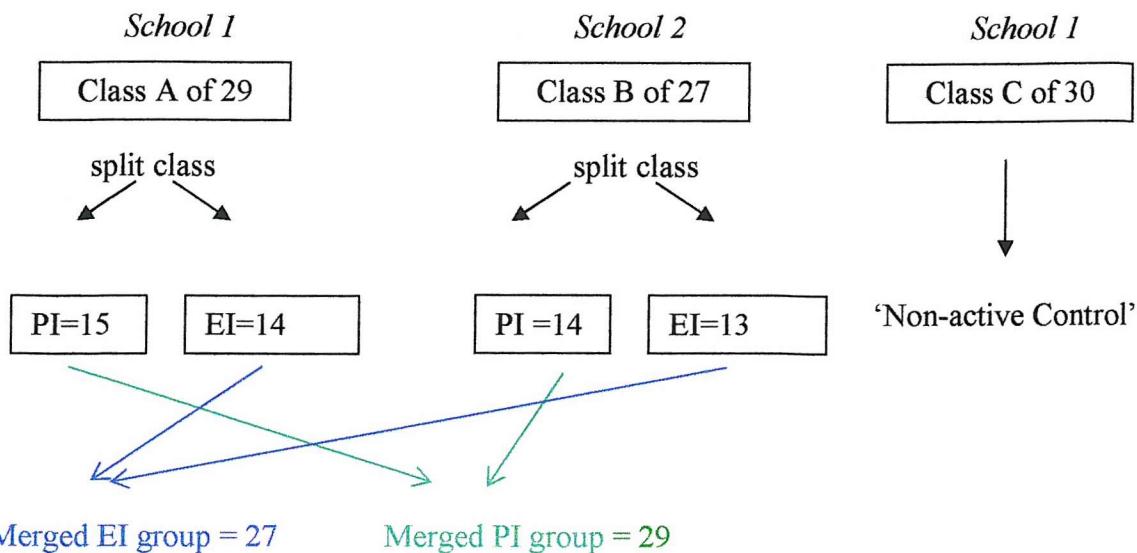


Figure 3.1: The Experimental Design.

29 PI and 27 EI pupils completed the study, though these were split between the two different schools. The number of participants was similar to the first PI study, VanPatten & Cadierno (1993): PI group ($n = 27$), Traditional Instruction group ($n = 26$), control ($n = 27$).

To divide classes A and B up into roughly comparable groups, randomised matched pairs were used¹⁶. The pre test scores were ranked and the pupil achieving the highest score was 'matched' with the pupil with the next highest score, and so on¹⁷. Each pupil in these matched pairs was then randomly assigned to a PI or EI group using Excel to generate random codes. The pupils were told that these groups had been formed because their averages in the tests had been exactly the same (there was some concern from the pupils that the groups were 'ability' groups). We referred to the PI group as the green group and the EI as the blue group. Each group was asked to move to another classroom the same number of times.

¹⁶ Henceforth, 'class' refers to intact classes, 'group' refers to one of the subsets of a class after it has been split, and 'merged group' refers to all the PI (or EI) learners taken from classes A and B.

Statistical analyses were carried out to ensure that the randomised matched pairs sampling had successfully generated groups with statistically equivalent scores at the outset. This was to eliminate the possibility that initial lack of parity could explain differences between groups at post and delayed post tests. Appendix 15 shows the results of these analyses for each measure: reading, listening, writing and speaking (see section 3.6 for a description and discussion of these measures and section 3.8 for a discussion of the statistical procedures used).

One analysis compared the PI and EI merged groups and class C at pretest using one way ANOVAs and post hoc tests, as in previous PI studies. It was found that the merged EI and PI groups had similar scores at the outset. However, class C had statistically significantly different pretest scores to the two merged instructional groups in the reading and writing measures but not in the speaking and listening tests. This issue will be addressed in chapters 5 and 6, as it suggests that treating class C as an equivalent 'instructional group' alongside the PI and EI merged groups will require special attention to ensure that any statistically significant impact of 'instructional group' on the test scores is not due to class C, but due to the experimental groups PI and EI.

The second analysis compared the PI and EI groups in each of classes A and B at pretest using independent samples t tests for equality of means. It was found that within each class there was no statistically significant difference at pre test between each instructional grouping for each of the measures.

These analyses have suggested that the randomised matched pair design resulted in statistically homogenous samples for learners experiencing EI and PI. This strengthens any claim in chapters 5 and 6 that any differences found involving the EI and PI groups at post and delayed post tests were due to factors other than differences in initial scores.

¹⁷ Scores used for the matched pairs procedure were: reading, listening and sentence level writing (a total of 150 test items). The speaking tests were excluded because they were not taken by all pupils; time constraints meant that the discourse level writing scores could not be included.

3.3.1 The use of two schools

Using two classes from different schools for the experimental intervention strengthened the reliability and validity of the study. Gall, Borg, & Gall (1996) suggest that the internal validity of an experiment in which intact groups are used can be maintained if there is more than one class per experimental treatment and the classes (though in this case, it was learners) are randomly assigned to the experimental and control groups (p. 490). Although the experimental classes were 'top set year 9' classes in similar schools, class A and B's mean achievement in the pretest scores used for the matched pairs was different (26.9% in class B and 43.5% in class A). This is analysed in detail in chapters 5 and 6. Using two classes therefore offered an opportunity to explore the effects of PI versus EI amongst pupils at slightly different developmental stages. Essentially, two 'mini studies' were available within each of which almost all learners had the same exposure to instruction outside the experimental intervention throughout the study and in the previous year.

In addition, amalgamating the PI and EI groups from the two schools to form 'merged groups' can be likened to using two similar 'mixed ability' classes. Furthermore, taking pupils from two classes meant that some account was taken of 'grouping effects' (or 'intraclass correlation') i.e. pupils in one class tend to have results more alike than if they had been the same students but scattered into different classrooms (Fitz-Gibbon 1996 p129).

3.3.2 The organisation of the teaching during the intervention

After the pre tests, class A (school 1) and class B (school 2) were divided into two groups for most of their lesson time for 6 or 7 weeks - one PI and one EI group. According to Doughty & Williams (1998) this is a relatively 'long' FonF intervention (p250). There were 12 'units' of experimental intervention materials, and, generally, each was taught in a different lesson. During the experimental intervention the PI and EI groups in both schools were video or audio recorded for every unit and any whole class, non-experimental teaching was also monitored (see appendix 10). Teachers A and B were given guidelines about how long each PI and EI activity should take and about how to carry out each activity, including how and when to give feedback to the pupils. The

class teacher and I alternated between teaching the PI and EI groups, to discourage the learners from associating one approach with one teacher - see appendix 14 (a & b) for details¹⁸. More details about the design of the teaching materials are given in subsequent sections (e.g. the language sample). An example of a PI and EI unit are provided in appendices 12 and 13.

During the same period, all lesson plans for class C were collected and two lessons were observed and recorded with teacher C.

3.3.3 The PI and EI Instructional Materials

As all previous PI studies have been carried out with older and more advanced learners it was important to adapt the design of the PI materials e.g. the complexity and content of the task and the range of vocabulary. Pilots of about one sixth of the draft PI and EI materials were carried out in four year 9 classes in two schools not participating in the main study. These pilots led me to reduce the EGIs to very short Overhead Transparencies (OHTs), gave me an impression of the timing of the activities and the suitability of the lexical items used and emphasised the importance of general issues such as creating enjoyable activities that 'made sense' and whose objectives were communicated to the pupils.

Both instructional packages (PI and EI) were balanced exactly for types and tokens of vocabulary and of the target structures. The number of illustrations used was kept as constant as possible. The length of the tasks (number of items and time required) was usually similar, though the nature of the referential activities did not always allow this. The reader is reminded that the main difference between the PI and EI materials was as follows: the referential activities in PI required the learners to attend to the meaning of verb inflections whereas all EI tasks¹⁹ required the learners to attend to the meaning of the lexical items or overall sentential meaning, thus any attention to the verb inflections would have been incidental to the task.

¹⁸ There were a couple of exceptions due to teacher absence but PI and EI pupils still had equal exposure to the same teachers.

¹⁹ and the affective PI tasks

Each unit generally had one or two 'topic areas' e.g. in the classroom, describing other people, the environment, describing towns, at the campsite, expressing opinions. All of these broad areas were included at some point on each school's scheme of work, although as the order of topics was different in each school, it was not possible to follow the sequence suggested by these schemes of works. Instead, the sequence of units was mainly driven by the target linguistic features. In most units, the teachers followed the experimental guidelines and informed the learners of the 'grammar' and 'topic/situation' objectives before each lesson, in line with recommended practice at both schools.

Appendix 16 gives an outline of sequence of linguistic features taught. Note that each unit focussed on one 'pair of form-meaning contrasts', though most of the inflections were included in more than one unit of work but 'paired' with different 'contrasting features'. As Doughty & Williams (1998) stress: "multiple encounters are required for engaging learning processes, such as noticing, hypothesis formation and testing, comparison and restructuring" (p253).

In line with previous PI studies and current IP theory²⁰ pair and group work was not promoted by either of the instructional types, except for about five of the final affective activities (PI and EI versions) where pupils could discuss the opinions they had expressed²¹.

3.4 The language sample for the teaching materials and tests

3.4.1 *The lexical items*

The vocabulary used in the instructional materials and tests was broadly taken from the schemes of work, worksheets and language used in the lessons prior to the intervention, and the textbook used by teachers A, C and D (McNab 1994). Each teaching activity included glossaries for vocabulary that the learners may not have known. The teachers

²⁰ VanPatten (2004) makes some attempt to incorporate a role for interaction (e.g. negotiation of meaning) in input processing in that it can free up working memory to allow detection of items in the input that may not have been attended to otherwise. However, this has not yet been developed further in IP or PI.

²¹ Learners were not required to produce the target features.

were also asked to check comprehension of items they thought their pupils may not understand.

3.4.2 *Choice of verb inflections*

Most of the verbs included in the teaching materials and tests were regular 'er' verbs as this represents 90 % of French verbs. However, some of the materials featured irregularities in present and perfect verb inflections. This was for several reasons: to support the notion that the instruction and tests were promoting / eliciting system (not item) learning; to maintain external validity (the schemes of work from both schools included a wide range of structures); in PI referential activities, certain irregular verbs were useful in terms of promoting the meaning for tense carried in auxiliaries (see section 2.2.3)²².

One short PI unit focused on the regular 'er' past participle, the referential activities forcing learners to interpret the meaning carried by the difference between the oral null / 'ons' inflections *versus* [e] and written forms er / e / ons / ez *versus* é. After this unit, the PI referential tasks focussed attention on the auxiliary as carrying meaning for tense.

Only reading activities were used to encourage learners to attend to written inflections which are not realised orally (e.g. 'ent' to convey 3rd person plurality, 's' to convey 2nd person singular). The materials and tests included verbs with obligatory contexts for 1st person plural and 2nd person reference, lacking in many studies to date (see appendix 6).

The teaching materials and tests promoted / elicited both simple and progressive functions of the French present tense. This increased the breadth of the target system under study, but it was beyond the scope of this thesis to analyse the acquisition of these different functions in the test results.

²² i.e. when the lexical verbs are homophonic (and very similar in written forms) in both tenses e.g. *je fais / j'ai fait, je dis/j'ai dit, je finis/j'ai fini* and the second person plural e.g. *vous mang[e] vous avez mang[e]*. The phonological similarity between *j'aime jouer* and *j'ai joué* also offered opportunities to promote comprehension of the auxiliary.

The range of ‘semantic types’ of verbs was intentionally broad (including, for example, state, activity, accomplishment and achievement verbs). This took some account of any possible bias caused by lexical semantics influencing the acquisition of tense and aspect (Anderson & Shirai 1994, Housen 1998). It was, however, beyond the scope of this thesis to investigate this issue in the test results.

3.4.3 *Creative construction?*

The sampling of language for the teaching and test materials had to provide opportunities for learners to engage in and demonstrate system learning (Harley 1994), rather than learn and recall items of routinised formulae which may over-represent learners’ underlying morphological knowledge. Both the materials and tests attempted to avoid the use / elicitation of routinised formulas common in teaching and interlanguage as documented in, for example, Myles, Hooper & Mitchell (1998) and Page (1999)²³.

3.5 Designing the tests

This section first discusses some general issues associated with the tests. In the following section, each test activity is described, with details of their objectives, linguistic content, timing, administration and scoring. Both sections are mainly descriptive, though some reference is made to general and SLA assessment literature and to FonF/S and PI studies.

3.5.1 *Low Stakes*

The learners were reminded before every battery of tests that results would not be passed on to their teachers or parents and that the main objective of the tests was to obtain a *class average* (so individual scores would not be highlighted). Low stakes tests were considered most suitable (contrary to Fitz-Gibbon 2000²⁴) for the following reasons: i) ethical considerations (learners were obliged to carry out the tests, with no warning, and did not receive feedback) ii) validity issues (low stakes tests could be used when the

²³ Although this was not always possible in a couple of items in the less controlled production tests, appendices 23 and 24 explain that certain ‘chunks’ were eliminated from the scoring.

teacher had not approved them and was therefore less able to 'teach to the test') iii) to lower affective filters and iv) to reduce any temptation to remember and learn test items. It was desirable to reduce the extent to which learned knowledge could influence scores as this is usually believed to reflect explicit knowledge, to be short-lived and not to represent the underlying developing system. Pupils were encouraged to do their best by appealing to their appreciation of the importance of research. In addition, it was likely that any lack of motivation would not adversely affect the experimental protocol as it would affect both PI and EI learners in similar ways. Pupils appeared enthusiastic and engaged throughout the tests, with one exception: during the written post tests at school 2 the pupils appeared unusually distracted as they were visibly and audibly annoyed at having to do a French test²⁵. Many learners appeared to 'rush through' the tests, in comparison with their behaviour in the pre and delayed post tests. This is discussed in the analysis of the written production results in chapter 6.

3.5.2 Task Familiarity

Skehan (1998) suggests task familiarity must be taken into account when designing tasks and tests as it can influence learners' performance. Previous PI studies have acknowledged that particular tests favoured either the PI (the interpretation tests) or the traditional groups (the production tests). This issue was less problematic in this study as both groups had input-based instruction. Care was taken to ensure that neither the PI or EI activities 'trained' the learners to do the tests. For example, the layout and the question format of the tests were significantly different from the teaching materials. The tests were designed to ensure 'world knowledge' did not influence the tests or the impact of the teaching materials, by making them as socially, culturally and gender-'appropriate' as possible (Bachman & Palmer 1996).

3.5.3 Different versions of the same test

Two versions of the same listening, reading, speaking and writing tests were used. This was to increase the sample of the language systems being tested and to reduce the

²⁴ who suggests that tests used for 'effectiveness' research should be curriculum-embedded, authentic and high stakes, in order that learners take the tests seriously.

likelihood that any improvement was due to learners becoming more familiar with particular test items. Half of each class took test 1 and half took test 2 for each administration of the test, except for the listening tests which were taken by all the learners in each intact class (though as class A had test 1 and class B had test 2, at the pretest, it was still the case that half of the scores for the 'merged PI and EI groups' were from test version 1 and half from version 2).

For the post tests the learners took the version of the test not taken at pre test. For the delayed post test they took the same version as in the pre test, but up to 6 months had elapsed between the pre and delayed post test.

To ensure equivalence of versions 1 and 2, the tests were piloted in 4 classes in three different schools (not those used in the main study). Each pen and paper test was taken by at least 20 learners and the two versions of the same test were compared. If items obtained different results by more than about 3 or 4 learners then the items were changed. Usually this entailed the use of a more common lexical item or reducing the complexity of the linguistic context for the target item. This piloting also led to refinement of the stylistic presentation of the tests so that they were identical.

A subsequent analysis of the actual test scores, using independent samples two tailed t tests²⁵, showed that the two versions of the tests had obtained statistically similar scores. The results are given in appendix 17. Any differences in scores found in the next chapters cannot, therefore, be explained by the use of two test versions. The only exception was the listening test. As this was administered to intact classes, these results are heavily influenced by a class and/or school effect (the same test version was taken by classes A and C and these classes achieved higher scores than class B at pre, post and delayed post tests, see chapters 5 & 6).

²⁵ The pupils were expecting a history lesson, but rearrangements were necessary due to timetabling changes.

²⁶ See section 3.8 for a description and discussion of the statistical techniques used. Parametric tests were used for parity with previous PI studies. However, as some of the data was not normally distributed

3.5.4 Task type

A meta-analysis of 49 FonF/S experiments (Norris & Ortega 2000) concluded that the impact of instruction may “be directly associated with the type of [test] response required from learners” (p486): studies using ‘selected-response’ and ‘constrained-constructed’ outcome measures had effect sizes between 0.38 and 0.91 standard deviation units higher than those using ‘metalinguistic judgements’ and ‘free-constructed’ response measures. VanPatten & Sanz (1995) also recommend that different modalities and task types be used to assess impacts of instruction, as they found that PI was more likely to have an impact on a written narrative task than on an equivalent oral narrative task (see also Sanz 1997). Skehan (1998) suggests a range of processing conditions (planned to unplanned, formal to informal, spoken and written modalities) in order to obtain a representative sample of different types of competence / performance.

A battery of tests was used in this study to elicit as comprehensive a range of language as possible, with a sample of different modes, modalities, levels of ‘spontaneity’ and text (word / sentence / narrative). The aim was to create broad measures, each comprising a range of task types (see section 3.6). Sentence *and* narrative level tests were used as IP theory is largely concerned with sentence processing, as emphasised by Sanz & VanPatten (1998), and as the PI framework has been criticised for its lack of use of less controlled (narrative) conditions. (Differential impacts of PI or EI according to task type are briefly analysed in section 6.2.2).

3.5.5 Post and delayed post tests

Figure 3.2 shows the timescale of the tests and experimental intervention.

	Pretests	Intervention	Post tests	Normal instruction	Delayed post tests
Duration	2 weeks	7 weeks	2 weeks	12 weeks	2 weeks
Cumulative timescale	2	9	11	23	25

Figure 3.2: The timescale of the experimental intervention

(possibly not checked in other studies) non-parametric tests were also used where necessary. Both the

Post tests were carried out over the two weeks immediately following the intervention i.e. assessing the impact of instruction that took place between 1 day and 9 weeks before.

The delayed post (dp) tests in this study took place approximately 12 to 16 weeks after the post tests, therefore assessing the impact of instruction between 14 and 24 weeks before. This is later than all PI studies published to date and most FonF/S studies reviewed in Norris & Ortega (2000) (where dp tests are usually administered a few weeks after the intervention). However, it is acknowledged that even dp test evidence may well be prone to backsliding and may not necessarily indicate acquisition in a 'permanent' sense²⁷.

3.5.6 Administration of the tests

I administered all the pen and paper tests (listening, reading and writing), all the sentence level oral tests and one third of the one to one oral tests (narratives and guided conversations). The remaining one to one oral tests were administered by my supervisor and a native speaker university lecturer, both accustomed to carrying out such elicitation procedures. They were both given detailed guidelines and met with me prior to the test to familiarise themselves with the tasks. Most pupils had the same interlocutor for all three tests, though this was not possible in 4 cases.

The battery of tests took place during a fortnight's lessons. Any potentially relevant prompts (e.g. verb paradigms) on the classroom walls were removed. In all three classes, the reading and listening activities were always carried out first. The second lesson in the 'test fortnight' was the 45 minutes writing test, though twice this had to be delayed until the end of the fortnight. The third lesson was used for the one to one oral tests. The remaining one to one oral tasks and the group orals (i.e. sentence level tasks) were carried out in the fourth lesson²⁸.

parametric and non-parametric tests produced similar findings.

²⁷ Certain interlanguage phenomena can be documented amongst learners with different amounts of exposure to the target language (see appendix 6), suggestive of re-structuring and backsliding.

²⁸ Pupils who missed tests did them either during the 3rd and 4th lessons, during lessons beyond the 'fortnight' or, on some occasions; pupils were willing to do tests during their breaks.

3.6 The language tests

Pre, post and dp tests were used to measure progression in the comprehension and production of French inflectional verb morphology in the present and perfect tenses. Test activities were grouped together according to their modality (oral / written) and mode (production / interpretation), in line with the research questions (section 2.5), so that there were 4 separate measures: listening, reading, speaking and writing. Most other PI studies have generally used just one short, assessment task in one modality for each mode (see appendix 4). However, in this study, claims regarding the results are strengthened as performance in both modalities was assessed and each measure is derived from more than one task, thus representing a wider sample of the language system under investigation and a wider range of performance types (see appendices 18 and 19 for breakdowns of the tests according to mode, modality, task type, verb type, person, number and tense). Tests used in previous PI studies informed the design of these tests (particularly the listening tests, as they were concerned with eliciting bottom-up processes, whereas most listening assessment literature focuses on the testing of 'top-down processes' (Brindely 1998)).

All learners took all of the tests, except the oral tests which were carried out with a subset of learners for practical reasons²⁹. Here, brief descriptions of the tests are given, structured according to mode and modality. The language sample in the tests was discussed in section 3.4. Fuller descriptions of the tests, including details of timings, distractor items, presentation, administration and rubrics are given in appendix 20. The actual tests (version 1) are provided in appendices 21 (listening and reading tests) and 22 (writing and speaking tests).

3.6.1 *Aural comprehension (listening)*

This was assessed in 4 activities, with 6, 10, 8 and 24 multiple choice items in each. The scoring was 1 or zero, out of a total of 48. Learners had to select the subject that was

²⁹ Overall, missing test data was: 1 class A PI pupil in the listening and reading pretests; 1 class B EI pupil in the writing pretest; 3 class C pupils in the writing post test; 2 class C learners in the writing delayed post test; 1 class C learner in the oral delayed post test.

appropriate for the verb they heard or decide whether the sentence they heard was in the past or present. The category 'not sure' was added in each activity to reduce the effect of the learners guessing (Alderson 2000 p217). In addition, statistical analyses were carried out to assess the likelihood that chance selection was playing a significant role in the results (see chapter 5).

3.6.2 Written comprehension (reading)

This was assessed in 6 activities with 8, 10, 10, 5, 5 and 17 multiple choice items in each. More response options were offered than items, so that learners using a process of elimination did not benefit towards the end as they 'use up' the options (Alderson 2000 p218). Scoring was 1 or zero out of a total of 55, excluding 8 distractors. Multiple choice formats where the same stem with different inflections is offered in a list format underneath the gap were not used as this would have forced learners to notice the different inflections. Instead, each activity was designed so that all the options offered (i.e. lexical verbs or subjects) were compatible on a 'semantic level' with any of the test sentences i.e. learners could operate at a semantic level of processing if they so wished, but every sentence would only be correct if learners demonstrated target-like interpretation of verb inflections. Pupils were told each time that they could ask me on an individual basis the English meaning of single French words and that, most of the time, I would be able to tell them. This was to reduce the effects of individual differences in vocabulary comprehension, even though the pilots had striven to eliminate such effects.

3.6.3 Written production

This was assessed via five tasks: three word/sentence level gap-fill tasks (28 sentences, with two points available for each verb) and two narrative level tasks (where the total was dependent on the obligatory contexts provided by each learner, see section 3.6.6). Before the sentence level tests learners had timed exposure to revision lists with English and French infinitive verbs. They could use an English verb prompt given at the end of each sentence to write the correct verb in the space provided. The narrative tasks required learners to describe what particular people do or did at the weekends or on holiday. The

scoring for both sentence and narrative level tasks was graded to allow for progression towards the target-like production (see section 3.6.5 and appendix 23).

3.6.4 Oral production

This was assessed using two types of tests: a) more controlled sentence level tasks and b) one to one narration and guided conversation tasks. The sampling procedure used to select 'representative learners' is described below. The scoring for both sentence and narrative level tasks was graded to allow for progression towards the target-like production (see section 3.6.5 and appendix 24).

a) Sentence level tasks

16 learners were selected from classes A and C and 17 from class B³⁰. This was done by selecting two boys and two girls (only in school 1, as all pupils were girls in school 2) from each quartile of the total reading, listening and sentence level writing pretest scores. As the pretest data was not available in sufficient time from class B, a score created from data provided by the school was used³¹. Equal numbers of pupils from PI, EI merged groups and class C completed the tests. Each pupil was given a tape recorder and lapel microphone and I directed groups of about 8 through the tasks.

The learners completed 3 sentence level semi-spontaneous oral production tasks, with a total of 27 sentences (with two points available for each sentence), where symbols had to be replaced with verbs inflected according to the context. Before two of these activities learners were given timed exposure to revision lists of the symbols used, the French verb infinitives with English translations, again to reduce the effect of familiarity with symbols and vocabulary, as in Benati (2001) and Houston (1997), amongst others.

³⁰ The difference in numbers was due to an administrative error, but the extra learner was retained in the analysis as it was deemed that more data constituted a stronger design than the slight difference in sample sizes, which could be accommodated by using suitable statistical procedures.

³¹ This score consisted of: English, Maths and Science National Curriculum levels at the end of year 8, the differences between pupils' chronological age and reading and spelling ages on entry to the school, 3 Cognitive Ability Test scores (verbal, quantitative, non-verbal) and an NFER Standard score (MALL). This data was collected by both schools for the 'Hampshire school data / value added scheme', deemed a reliable system of its type (Goldstein 2001 p440). Although these scores are not based on foreign language learning, they are from standardised tests that give a normal distribution of scores.

b) One to one narrative and guided conversation oral production

Narrative level semi-spontaneous oral production has only been elicited in one previous PI study (VanPatten & Sanz 1995), though insufficient tokens of the target form were elicited, as acknowledged by the researchers. Skehan (1998) points out the difficulty of trapping forms in tests whilst also maintaining spontaneity. In this study, one to one oral production tasks included two narration tasks and two prompted conversations (e.g. 'normalement le weekend, nous...'). Although the latter perhaps elicited less spontaneous production, they were deemed necessary to elicit first and second person plural contexts, particularly as other studies with similar learners have not elicited significant numbers of these and, given their distinctive phonemes, they are important for documenting an emerging inflectional system.

The one to one oral tests adapted some features of those used in Mitchell & Dickson (1997), Page (1999) and VanPatten & Sanz (1995). For both the narration tasks, learners heard a 'model' response to activate recall of lexical items and to lower affective filters. The risk of learners being able to 'imitate' target-like morphosyntax was considered to be minimal, given the use of such practice in other studies (Mitchell & Dickson 1997), the reliance on elicited imitation procedures in SLA research (Henning 1983, Naiman 1974, Myles 1995) and cognitive psychology studies which show the limitations of short term phonological memory (Papagno, Valentine & Baddeley 1991).

Scoring of these tasks was done, as in other PI studies, out of totals based on the obligatory contexts³² produced by each learner.

From each class, eight of the learners who had taken the sentence level task (4 PI and 4 EI learners from classes A and B) undertook these semi-spontaneous tasks (two pupils from each quartile: a boy and a girl from the classes in school 1, 2 girls from school 2). Their scores were amalgamated with the sentence level tasks to make one 'oral measure'. The numbers from each sample of learners taking each type of test are given in table 3.1.

³² See section 3.6.6 for a definition of 'obligatory contexts'.

		<i>n</i> level	Sentence	<i>n</i> Narratives / conversation	/
Class A	EI	8	4		
	PI	8	4		
Class B	EI	7	4		
	PI	10	4		
Class C		16	8		

Table 3.1 Numbers from each sample taking oral test tasks

Chapter 6 presents analyses to show that there was no statistically significant difference between the results obtained by the sentence level and narrative level elicitation techniques and that learners who undertook both the sentence and narrative level tasks and those who undertook just the sentence level tasks were from the same population i.e. that although the spoken production score was derived from two task types for some learners but from one for other learners, it can be considered a valid measurement.

3.6.5 Measuring interlanguage in the production tasks

The interlanguage³³ of this level of learner (appendix 6) suggests that accuracy is often not a sensitive enough measure to reveal progression e.g. because target-like non-routinised oral production of the perfect tense is unlikely. Skehan (1996) and Doughty & Williams (1998) both argue that the emergence of many non-target IL forms often reflects increasing complexity in the interlanguage system. Intermediate effects, such as increased use of inflections (regardless of their target-likeness) must also be measured. Many studies have documented advances in IL development as an increase in attempts at the target form, for example, Day & Shapson (1991), Williams & Evans (1998), Doughty & Varela (1998), Leeman et al. (1995). Several PI studies have a 'graded scoring system', sensitive to partial effects of instruction: 1, 0.5 or 0 points depending on the degree of accuracy e.g. VanPatten & Cadierno (1993), Benati (2001) and Cadierno (1995) (the latter two involving verb inflections for tense, person and number). Sanz & VanPatten (1997) criticise Salaberry (1997) for using an 'all or nothing' type scoring system, where all errors received -1.

³³ For scoring it is considered that IP theory is only testable if the steadiness of IL (rather than unconstrained variation as in Tarone 1988) is assumed to hold at the time of testing within learners, as in other FonF/S and PI studies.

The graded scoring procedures used in this study are detailed in appendices 23 and 24. They are summarised as follows: 2 points were awarded for a target-like inflected verb, 1 point for a verb with a non-target like inflection and 0 points for an infinitive or no verb. For perfect tense inflections, half points were awarded for non-target like attempts at auxiliaries and, in written production, past participles.

To improve the reliability of the scoring, a few weeks after the data had been scored, tests from four learners were re-scored and the results were almost identical. In addition, my supervisor was consulted on some scoring decisions. Although it would have been preferable to obtain inter-rater reliability, this was beyond the means of the current study.

3.6.6. Defining obligatory contexts

As in most FonF/S and PI studies (e.g. VanPatten & Sanz 1995), the scoring for this study was based on use of forms in obligatory contexts for narrative and guided conversation oral and written tasks. Obligatory contexts did not include utterances (written or oral) that did not have a verb or a subject, were inaudible or were repeated verb types with same subject (there were very few examples of this as learners were asked to use a different verb for each utterance / sentence). In writing, if there was a full stop, the subject had to be re-stated³⁴.

3.7 Monitoring the intervention process

As discussed in section 3.1, monitoring the implementation of the experimental protocol and participants' attitudes towards it was essential for the purposes of validity and reliability.

3.7.1 The operationalisation of the experimental materials

Measures were taken to ensure that the teachers conducted similar and valid operationalisations of PI and EI. Three PI researchers (including VanPatten) looked at some of the PI teaching materials and gave me some advice about their design. The

teachers were given fairly rigid guidelines (shown in appendices 12 and 13) about how to direct the activities. In addition, teachers A and B were asked to watch a video of me carrying out the first PI unit. Teacher A commented on the video, showing that she had watched it. I watched videos of both teachers teaching the second unit of PI and made a couple of comments to each teacher regarding the videos. After the intervention, a sample of four video or audio recordings of each teacher carrying out the PI and EI materials was analysed (see appendix 14 for a record of the protocol and evidence collected). I was satisfied that teachers A, B and I had met the criteria to classify the instruction as PI and EI. Furthermore, any possible effects of teachers' subverting the protocol were reduced by the fact that half of each package was delivered by me³⁵. Unfortunately, inter-rater reliability was beyond the means of the current study.

3.7.2 *Pupils' attitudes & opinions*

Likert ratings and a questionnaire were used to make a very general assessment of the 'novelty effect' of the intervention. If *both* PI and EI groups demonstrated a positive attitude towards the materials then this would help to support the claim that any differential improvement was probably due to the specific nature of the different task types. If *only* the PI group showed a positive attitude *and* greater linguistic progress then the improved learning may not necessarily be due to the usefulness of the pedagogy in terms of its underlying theories, but perhaps due to increased motivation (or the increased motivation may be due to a sense of achievement)³⁶. Clearly this in itself would be a useful finding, given that motivation is a major challenge facing MFL teaching in the UK.

a) *Likert questions*

Two Likert rating scale questions (appendix 25) about learners' perception of difficulty and preference of French compared to other school subjects were filled in three times: before the pre, post and dp tests. These are analysed in chapter 4.

³⁴ See appendix 23 for further details regarding obligatory contexts in the narrative writing tests.

³⁵ In addition, two likely reasons teachers may have had for subverting the protocol (their deductive approach and absence of production practice) were equally present in both PI and EI.

³⁶ No specific motivation research framework is adopted here. It is emphasised that the data can only be used as a fairly crude indication of attitudes and opinions.



b) Questionnaire

Pupils in classes A and B completed a questionnaire eliciting their opinions of the experimental intervention materials during the same lesson as their last unit of work (appendix 26). It is acknowledged that the data will probably be affected by the fact that I was involved in the teaching and also in handing out the questionnaires. The questionnaires were not filled in anonymously as I wanted to match up comments with individuals' scores, attendance and behaviour. However, I reassured the pupils that I wanted them to be very honest with me, that I had no influence over their school assessment, that their class teacher would not see their responses and that any use I made of them would be anonymous. Furthermore, any possible pollution of the data is likely to have affected both PI and EI learners equally and so should not affect the interpretation of the results of the questionnaire. It is acknowledged that pupils' enjoyment of the activities may have been influenced by the fact that they were aware of their involvement in an experiment (the Hawthorne effect). Again, this is likely to have affected both groups in similar ways. In addition, the questionnaires included questions to elicit a self-report on whether the learners thought the lessons *had* been different to their normal lessons. This data is also analysed in chapter 4.

c) Focus Groups

Several pupil focus groups were also carried out, towards the end of the intervention, regarding the pupils' feelings about the PI and EI materials. 12 PI pupils and 10 EI pupils, in equal numbers from each school, participated in the discussions in groups of 2 to 4. The results generally supported the findings from the questionnaires. A few issues raised in these discussions are mentioned in subsequent chapters. However, a detailed analysis of this data is not included in the thesis, partly because the self-selection procedure and my direct involvement in both the focus groups and the teaching may have affected the reliability of the comments.

3.7.3 Interviews with teachers A and B

These were semi-structured audio-recorded interviews for about one hour (see appendix 30). The interview with teacher A was almost 3 weeks after the post tests and with

teacher B it was almost immediately after the post tests. Samples of the teaching materials were used in the interview to prompt discussion about the aims, strengths and weaknesses of each activity type. Three points from this data relevant to this thesis are discussed in chapter 4³⁷.

3.7.4 *Informal researcher's diary*

I wrote a diary noting a range of issues (frustrations, problems, successes) including comments from teachers and pupils that were not recorded, one of the advantages of participant observation (Griffiths 1998 p40). The only significant use of the data from this diary in this thesis is to provide evidence of pupils' distress during the writing post tests in class B, as described in section 3.5.1 and in chapter 6.

3.8 Procedures for the analysis of the achievement tests

A range of analysis techniques was used to check for differences between the test scores in the PI and EI groups and in classes A, B and C. Literature used regarding statistical techniques included: Jones & Bryant (1998), Diamond & Jeffries (2001), Cohen, Manion & Morrison (2000), Woods, Fletcher & Hughes (1989), Field (2000), Kanji (1993), Holmes (2000) and Bryman & Cramer (1997). Some aspects of the statistical analyses were checked by two university lecturers who specialise in statistics for the social sciences. Detailed mathematical discussion of the statistical procedures is beyond the scope of this thesis and it is emphasised that procedures used are common in social science, FonF/S and PI research. The key statistical procedures used in the following chapters are described and their use justified here. A few procedures that are less recurrent in the thesis are described as they are used in chapters 4, 5 and 6. This section discusses the following issues: creating measures of language ability, eliminating data, choosing parametric or non-parametric tests, testing the normality of distributions, ANOVA (including single and repeated measures, planned contrasts and post hoc tests), equivalent non-parametric tests and effect sizes.

³⁷ Data was also collected regarding the impact on teachers and pupils of carrying out a quasi experiment in an educational setting, although it was beyond the scope of this thesis to analyse this data.

3.8.1 *Eliminating data*

Some studies in the FonF research tradition eliminate data that are considered to be 'outliers' (e.g. if they render the data not normally distributed). In addition, PI studies eliminate learners' scores on the grounds that their pretest results were 'too high to show improvement'. However, this removal of data is contentious with small sample sizes (e.g. apparent outliers in small data sets may actually be well represented in the population as a whole). In addition, as Salaberry (1998) points out, there is inconsistency in the cut off point used in PI studies (despite the fact that the same tests and educational level of learners was used as in the original study, VanPatten & Cadierno 1993). For the purposes of this study it was considered preferable to include all data points in the final analyses, particularly as one of the research questions was to investigate the generalizability of the effects of PI and EI between two classes at the same general educational level. In addition, the observation, interviews, questionnaires and researcher's diary helped to provide some richer data capable of explaining some variation if necessary.

3.8.2 *Choosing parametric or non-parametric statistical tests*

Parametric tests are usually considered to be more powerful than non-parametric tests (i.e. they can detect more 'differences', Bryman & Cramer 1997, Field 2000) and most FonF/S and PI studies use parametric tests. There is therefore a preference to present the results of parametric tests in this thesis. However, data must usually satisfy several criteria if parametric tests are to be used, involving issues related to data type, sample size, distribution and variance.

The achievement tests produce interval data and this is the right type for parametric tests.

Several guidelines exist on sample sizes necessary in order to carry out parametric inferential statistics. Diamond & Jeffries (2001) recommend a sample size of about 10 or more if the population is normally distributed, although Bryman & Cramer (1997) recommend non-parametric tests for samples under 15 (p118). A sample of 25 or more for parametric tests is suggested if the population is skewed (Woods, Fletcher & Hughes 1989). Goodwin & Goodwin (1996) suggest that for every 'predictor', 10 subjects are

needed (p41). In this study $n=86$ and three predictors will be used in final analyses: time of test, type of instruction and class. Overall, it seems that the sample sizes are appropriate for parametric tests, including for the data from instructional groups within classes ($n = \text{about } 15$) (though the oral and combined production scores are possible exceptions, described in sections 3.6 and chapter 6).

It has been suggested that parametric tests should only be used on data which also satisfy the following criteria: the distributions of the samples are normal and the variances of all variables are homogenous. Therefore, each of the variables for each of the samples is checked for these assumptions. The outcome of these checks is not always strictly respected i.e. parametric tests are presented where some of the assumptions were not upheld consistently. This is for three main reasons, as follows.

First, there is a preference for parametric tests as they are considered to be more powerful than non-parametric tests. Parametric tests give information about the size of the differences between scores by using the mean, standard deviation and variances of the group scores (not just the rank of each score) and are therefore more likely to detect a significant difference where there is one.

Second, it has been argued that it is not necessarily vital that data meet these assumptions. Bryman & Cramer (1997) suggest that there is recognition that some variables in the social sciences do not exhibit the characteristics of a normal curve and therefore variables have to be treated as though they were normally distributed (p96). They also review the statistical literature which suggests that these criteria do not have to be met for many parametric tests to remain valid (pp117-8). Moore & McCabe (1989) maintain that the ANOVA F test, used in this study, "is relatively insensitive to moderate nonnormality and unequal variances" (p732). Field (2000) also suggests that most multiple comparison procedures perform relatively well under small deviations from normality. However, he points out that some can perform badly when group sizes are unequal and when population variances are different but gives a range of possible post hoc tests that have been developed with these assumption violations in mind. Overall, the

literature suggests that parametric tests can be used with caution when some of the assumptions are not entirely upheld.

Third, all previous PI studies have used parametric tests and they are therefore used in this thesis in order to maintain parity with other studies. (In fact, only one PI study (Sanz & VanPatten 1998) has discussed the need for normally distributed data, and it is possible that this issue has been neglected by most PI and FonF studies (e.g. see Allen 2000 p81.))

However, if the assumptions necessary for particular parametric tests are not upheld, non-parametric tests were also used (as recommended in Bryman & Cramer 1997 p118). Their results were usually in line with the parametric tests, in terms of statistically significant findings, and therefore they are not presented in the body of the thesis but are provided in appendix 27.

3.8.3 Testing the normality of distributions

Although many parametric tests stand up well to small deviations in normality, some caution is required if the data are heavily skewed. For each measure, the normality of the distribution of the scores from each sample (PI, EI, classes A, B and C) was assessed using the Kolmogorov-Smirnov (K-S) test (and the Shapiro-Wilk statistic for borderline cases and small n, as suggested by Field 2000 p46) to test the similarity of the samples to a normal distribution³⁸. The results of these tests are shown in appendix 29 and are summarised in the body of the thesis before the analysis of each measure. It is acknowledged that the calculation of the K-S statistic relies on the sample data (i.e. the observed means and standard distributions, rather than data from a known population) and that this can be regarded as unhelpful as the observed values are the very values that are being assessed for normality. However, the K-S test is a widely-used and recommended statistical procedure (Field 2000 p46) and is therefore considered sufficient for this study.

³⁸ As repeated measures ANOVAs, one of the principal analysis procedures used in this thesis and in PI studies, exclude missing cases listwise (i.e. learners' scores are excluded if they missed any one of the tests), tests of normality carried out in this study also excluded cases in this way.

P values above 0.05 indicated that there was no significant deviation from normality at the 95% confidence level. However, if p values were above 0.01 it can be considered that the data are not significantly different to a normal distribution at the 99% confidence level³⁹. Increasing the confidence level makes it easier to assert that there is no difference between the observed distribution and a normal distribution (i.e. to accept the null hypothesis). However, this clearly increases the chances of making a type II error i.e. accepting the null hypothesis when it is false. A 99% confidence level can only therefore be used to assert borderline normality.

3.8.4 Analysis of Variance (ANOVA) and paired comparisons

ANOVAs are suitable for studies where more than two means are tested (pre, post and dp tests). They are also appropriate because there were two independent variables each with more than two 'levels': instructional type (PI, EI and the non-active control) and class (A, B and C). A repeated measures (two-way) ANOVA was used because there were 3 'repeated' measures (pre, post and dp tests) for each learner (the few exceptions to this were treated as 'missing data' and excluded). Although individual related t tests could be used to compare the mean difference between pairs of scores, this is not valid if used repeatedly on the same data set, as there is increased risk of making Type 1 errors (i.e. rejecting the null hypothesis of no difference when in fact it should be accepted⁴⁰). ANOVA procedures compensate for this before paired comparisons can be carried out.

The ANOVA test produces an F-ratio which compares an estimate of the between-groups variance⁴¹ with an estimate of the within-groups variance by dividing the former by the latter. When the F-ratio is higher (because the estimated variance between the groups is considerably higher than the variance within the groups) then the differences between the means of two or more groups is unlikely to be due to chance. To assess whether the F

³⁹ This is because increasing the level of statistical confidence level before accepting the null hypothesis (i.e. that there is no difference between the observed data and a normal distribution) reduces the possibility of rejecting this null hypothesis when it is really true.

⁴⁰ Repeated paired comparisons increase the error rate across statistical tests conducted on the same experimental data due to the fact that the 95% chance of not making a type 1 error is multiplied over several tests, which increases the chance of making the error (see Field 2000 p243-4).

⁴¹ Variance shows the dispersion of data around the mean and is the square of the standard deviation.

value is 'high' it is compared to the critical value of the F statistic which depends on the total n, the number of dependent variables and the desired confidence level of wrongly rejecting the null hypothesis. The F statistic gives an indication of whether there is any difference between *any* of the various groups of data.

Planned contrasts (Field 2000 pp258-271) were then carried out to see *where* this difference lay in terms of the instruction type (PI, EI, class C) and class (A, B, C). Previous PI studies have relied on post hoc tests, rather than planned contrasts. Planned contrasts are distinct from post hoc tests in that they test directional hypotheses (e.g. that the post test scores will be higher than the pre test scores) rather than non-directional hypotheses (e.g. that the post test scores will be different to the pre test scores). Another advantage of planned contrasts is that they facilitate paired comparisons on interaction effects (e.g. to test whether instructional type AND class had an impact on scores at post test compared to pre test). Previous studies have not had this 'class' variable, each study being carried out at just one institution. Planned contrasts also maximise the use of repeated measures (i.e. the analysis of *progression* between each test). In previous PI studies two separate *one way* ANOVAs and post hoc tests were required on the post and dp test scores to analyse whether there were differences between instructional groups *at* each of the tests (as post hoc tests can not compare the impact of 'instructional type' on post test to its impact on pre test). Such a procedure runs the risk that any initial differences between groups, although not statistically significant at pretest, may still render differences between groups at post test and dp test statistically significant.

In this study, two within-subject planned contrasts were carried out between each pair of dependent variables for every measure: a simple contrast (where the post and dp test measures were each compared with the pre test) and a repeated contrast (where each test was compared with the last i.e. post to pre, dp to post⁴²).

One way ANOVAs were carried out to compare the PI and EI samples at pre test (section 3.3.1) and the classes at pre test (section 4.2). One way tests were suitable for these

analyses as there were no repeated measures but there were 2+ *unrelated* groups. For these, a variety of post hoc tests were used according to, for example, whether the sample sizes and variances were equal (Field 2000 p276).

Each time an ANOVA was carried out certain assumptions were checked as follows.

First, the normality of the distribution was assessed using K-S tests (described in section 3.8.4)⁴³.

Second, the homogeneity of variance of the between-subjects effects (i.e. the test scores in the different groups of instructional type and class) was checked using Levene's test for homogeneity of variances (available on SPSS, Field 2000 p238). The results of these tests are shown in appendix 32 and are summarised in the relevant sections. To rectify differences between variances the data can be transformed. Field (2000) (p284 & p365) and Howell (1997) (p323-329) recommend the widely-used procedure of using the square root of the data (thus making the range of data more 'compact') and this was used in this study where necessary. However, if the results of the analysis using the transformed data were no different (in terms of statistical significance) to that of the non-transformed data, then the results of the analysis of the non-transformed data are presented. This was because transformed data can be difficult to interpret (e.g. the impact of instruction is on 'the square root' of the test results) and there is no tradition of using transformations within FonF/S research. For the same reasons, other types of transformations have not been attempted in this thesis.

Third, for repeated measures ANOVAs it was also checked that there was no sphericity in the data, using Mauchly's test of sphericity (also available in SPSS, Field 2000 p333)⁴⁴. This assumption requires that the variances of the *differences* between the repeated measures are homogenous. If this is not upheld, the test can lose power and the

⁴² Though clearly the contrast of the post to pre tests was a repetition of the simple contrast carried out for this comparison.

⁴³ Transforming the data using the formula used in VanPatten & Sanz (1995) for the production data scores: $y' = 2\arcsin(\sqrt{y})$, did not render the data any more normally distributed.

F statistic may not be valid. However, 'corrections' are produced by SPSS to overcome a violation of this assumption, by changing the degrees of freedom (used to compare the observed F-ratio against the *F*-distribution). These corrections are widely used and considered a reliable procedure in social statistics (Field 2000 p333-4). Field recommends the Greenhouse-Geisser correction (if none of the corrections make a difference to statistical significance) as this is the most conservative of the estimates of sphericity having the tightest hold over type 1 errors (where the null hypothesis is wrongly rejected). Results of Mauchly's tests are given in footnotes at the relevant points in chapters 5 and 6.

3.8.5 Equivalent non-parametric tests

When the procedures above suggested violations of the relevant assumptions, equivalent non-parametric tests were carried out in addition to the parametric tests as they are 'assumption free' in that they do not require normally distributed data and can be used with ranked data. Usually the results did not contradict the parametric tests, and so they are presented in appendix 27. Otherwise, they are discussed in the relevant section.

The equivalent procedure to the repeated measures ANOVA is the Friedman test for finding differences between 3+ related samples (Siegel 1956 pp166-172, Bryman & Cramer 1997 p138). Friedman (1937) (in Siegel 1956) showed that there was little or no loss of power with this statistical test when compared with the F test (the ANOVA). Siegel (1956) writes "it would be difficult or even impossible to say which is the more powerful test" p172. The Friedman test ranks all the scores from all the repeated measures. If there are statistically significant differences between the total ranks within each factor (i.e. pre, post or dp test) then this indicates that one or more of the factors had some impact on the scores. Once the Friedman test has found some statistically significant difference, then comparisons can be carried out on pairs of the variables (e.g. pre and post) - this helps to control for the fact that simply carrying out multiple paired comparisons would increase the likelihood of finding a statistically significant difference where there is none (see section 3.8.5).

⁴⁴ For both Levene's and Mauchly's tests the p values should be above 0.05 in order to accept that there

These tests were done in several 'layers', as interaction effects are not available in non-parametric tests. Once the Friedman test found a significant difference, then Wilcoxon matched pairs signed ranks tests were carried out (Jones & Bryant 1998 pp67-70, Bryman & Cramer 1997 pp138-9, Siegel 1956 pp75-83). This test can take account of both the direction and the magnitude of differences between pairs of related scores - the non-parametric equivalent of the t test - by ranking them and then summing those with the same sign (i.e. direction of difference). One tailed tests were carried out to test the hypothesis that the post and dp test scores were statistically significantly higher than the pre test scores, to ascertain whether the different instructional types and classes had different effects on the scores. As there was no experimental intervention between the post and dp tests, two tailed tests were carried out to test the null hypothesis that there was no statistically significant difference between these tests.

The Kruskall Wallis H test was used for comparing 2+ unrelated samples (the non-parametric equivalent of the one-way ANOVA test), and is necessary to compensate for the increase in the probability of Type 1 errors involved in carrying out the subsequent multiple paired comparisons (Siegel 1956 pp184-193). Once the Kruskall Wallis test found a statistically significant difference, the Mann-Whitney U test was used to compare the mean ranks of pairs of unrelated samples (Field 2000 pp49-54, Jones & Bryant 1998 pp65-70). According to Bryman & Cramer (1997) the Mann-Whitney test is about 95% as powerful as the t test (i.e. the t test requires 5% fewer subjects to reject a false null hypothesis) (p145).

3.8.6 Effect size

Norris & Ortega (2000), amongst other social scientists (e.g. Fitz-Gibbon 2001, Rosenthal 1991, Glass, McGaw & Smith 1981) recommend using effect size as an indication of the magnitude of the effect of interventions. Cohen, Manion & Morrison (2000) and Hunter, Schmidt & Jackson (1982) provide introductions to effect sizes and

were no significant differences between the variances.

their uses. They are a calculation of the difference in means between two groups in terms of the standard deviation.

$$\text{Effect size} = \frac{(\text{mean of experimental group} - \text{mean of comparison group})}{\text{pooled between-groups standard deviation}}$$

This gives some indication of the size of any difference in impact of interventions, which, according to Fitz-Gibbon (1984) is often a more useful indicator in any evaluation of educational practices as it does not rely on some 'arbitrary' level of statistical significance. In addition, the calculation of the effect size is essential if small scale studies are to contribute to a growing body of knowledge which can be synthesized using meta-analytic techniques e.g. Norris & Ortega (2000) (discussed in section 3.1.3).

To date, Collentine (2004) is the only PI study to use this technique. Collentine uses the method of calculating effect size recommended in Norris & Ortega 2000 i.e. Cohen's d (Cohen 1977). In this study, in line with Norris & Ortega (2000) and Collentine (2004), the pooled standard deviation was used as the denominator in the calculation of Cohen's d ⁴⁵. This reduces some of the sampling error due to the relatively small sample size and lack of random selection (Norris & Ortega 2000 p443, Fitz-Gibbon 1984 p139).

Following Norris & Ortega's (2000) adaptation of Hunter & Schmidt (1990 p271):

$$\text{Pooled standard deviation} = \frac{(n_1-1)S_1 + (n_2-1)S_2}{(n_1-1) + (n_2-1)}$$

where n is the sample size of either group and S is the standard deviation of either group.

The work of Norris & Ortega (2000) was used to establish 'benchmarks' on effect size for this study. They found that L2 instruction in particular language forms induced target-oriented gains of 0.96 standard deviation units (sdu) compared to a control group which did not have any exposure to the target form. When the control/comparison group was exposed to and interacted with materials in which the L2 forms were embedded they

⁴⁵ (as opposed to the control group's standard deviation, as recommended in Fitz-Gibbon 2001 and Glass, McGaw & Smith 1981).

found a mean effect size of 0.75 sdu. They conclude that these effects are "probabilistically rare" (p500).

In this study, in order to eliminate the impact of differences between groups at pre test on effect sizes at post and dp tests (the comparison groups did not begin at 'exactly' the same baseline), the gains⁴⁶ between the tests were used to calculate the effect size between particular pairs of groups i.e.:

$$\frac{(\text{mean gain of PI} - \text{mean gain of EI})}{\text{pooled between-groups standard deviation of gain}}$$

The effect sizes of pre-post, of post-dp and of pre-dp test gains is presented for each measure. Three pairs of experimental groups were compared: the merged PI and EI groups and the PI and EI groups in classes A and B. In addition, the effect sizes of gains between pre and post tests in class A compared to class C were also calculated for reasons presented in chapters 5 and 6.

3.9 Summary & original features of the study

This chapter began with a review of carrying out educational experiments. It argued for a mixed-mode approach to carrying out experiments, including the analysis of process and attitude data. The second part of this chapter described the data collection procedures, the design and implementation of the quasi-experiment and the analysis procedures used.

Drawing on chapters 1, 2 and 3, the first summary below (section 3.9.1) suggests features that have *improved* upon PI (and FonF/S) research to date in educational settings in terms of its generalisability, reliability and validity. The second summary (section 3.9.2) suggests how this study has made a *unique* contribution to the PI and IP research agenda.

⁴⁶ There were a very few occasions where the difference between tests was not actually a gain but a slight loss, but for the sake of consistency and clarity the 'difference between tests' is referred to as 'gains'.

3.9.1 Features which have improved the PI (and FonF/S) research tradition

- PI had not previously been investigated in UK secondary schools
- The learners are younger and are at an earlier stage of linguistic development than in other PI studies.
- The experimental exposure was longer, and over a more extended period of time, than in other PI studies, more reflective of normal educational practice.
- The range of linguistic features is broader than in previous studies, again more reflective of normal educational practice.
- This is the first study to explore the impact of PI on French verb inflections. Other PI studies using French have looked at syntactic features. The only other PI research looking at verb inflections was a study of the (less redundant) Italian future and Spanish preterit tenses.
- The time between the experimental intervention and the post and dp tests was longer than in other PI studies, and than in many other FonF/S studies.
- The materials ensured that in the PI referential activities one language feature was compared with another to draw learners' attention to the meaning carried by this difference in form (several PI studies have failed to do this).
- Satisfactory numbers of target feature tokens were elicited in semi-spontaneous oral production (unlike VanPatten & Sanz 1995 where, in any case, the comparison group had no exposure to the target feature).
- The effect size was calculated for each measure.

3.9.2 Features unique to this study

- PI was compared to another input-based approach, which, amongst other objectives, will allow exploration of the principles of IP as operationalised by PI i.e. 'forced' detection of target features compared to 'incidental' detection.
- Carrying out the study in two different classes (in different schools) may help to develop our understanding of the role of different developmental stages in the processing of input.

- Both groups had identical EGI, reducing the number of potentially extraneous variables.
- Broad measures of language development were taken which included a range of performance conditions (e.g. different modes, modalities, levels of control).
- Data regarding the affective impact of the materials on teachers and learners has been collected.

Chapter 4 The Learning Contexts and Participants' Attitudes

This chapter contains analyses of the data collected on the learning contexts and the attitudes of the learners and teachers to the intervention materials and French in general. Research question 4 is: "Do the same results obtain in learners from two similar classes from different schools?". Therefore, it was necessary to explore how the two experimental classes and the non-active control class compared to each other across a range of issues to assess the validity of the study e.g. to what extent was the experimental protocol adhered to?, what was the extent and nature of contextual variation? More specifically, as it transpired that PI and EI had different impacts in each class (see chapters 5 and 6), it was of relevance to know whether these classes were substantially different in certain key aspects outside and within the experimental protocol e.g. general organisation and approach to the non-experimental teaching and grammar pedagogy, learners' interpretation and production of the target verb inflections at the outset, learners' and teachers' attitude to French and the experimental teaching materials.

An analysis of the classroom contexts is provided in section 4.1. It presents 'portraits' of the three classes involved in the study, before, during and after the intervention, as background to interpretation of the test results in chapters 5, 6 and 7. Section 4.2 presents statistical analyses of the pre test scores, in each of the measures, examining the comparability of the three classes at the outset¹. Section 4.3 presents analyses of the attitude data i.e. the scores obtained from the Likert scale questions (completed at pre, post and delayed post test) and the questionnaire data (completed by EI and PI learners after the experimental intervention). This data was required to assess whether one of the instructional types was favoured by the learners and/or teachers as this can affect the validity and reliability of findings from educational experiments (see section 3.1.2).

¹ Section 3.3 and appendix 15 presents analysis showing that the experimental groups (merged PI and EI groups and the PI and EI groups in classes A and B) were statistically similar at the outset.

4.1 Portraits of the lessons

This section presents a portrait of the lessons in each of the three classes prior to, during and after the intervention. Descriptions of the actual experimental intervention activities (EI and PI) are very brief, as they were described in detail in chapters 2 and 3 and examples can be seen in appendices 12 and 13. Before these portraits are given, as many of these issues could be linked with departmental and whole-school policies and practices, there is a brief description of the environment outside the classroom in terms of matters such as schemes of work, departmental co-ordination and whole school policies regarding MFL learning in each school.

The portraits are based on detailed observers' notes (for example see appendix 11), audio and some video recordings. Appendices 10 and 14 give details of the evidence collected. In addition, the diary I wrote during the study and the interview with teacher B at the start of the academic year have provided further material for these descriptions.

These portraits give a general picture of how the three classes differed from each other, how the planned intervention compared to their normal instruction and the extent to which the target language features of this study figured in the lessons. An important aim was to indicate the expectations learners in each class probably had regarding the content, structure, demands and sequence of activities, and, in particular, their expectations of how formal accuracy is approached in MFL learning. The portraits are used in later chapters to suggest that contextual characteristics may offer some explanation of why the test results presented different patterns in the two different schools.

These aims motivated the choice of issues discussed: 'general classroom management', 'target language', 'error correction', 'resources & activity types' and 'focus on the forms targeted by the experimental intervention'.

These issues have their roots in both general 'effective teaching' literature and

'FonF/S²' frameworks, though reference to specific frameworks or theories is not made as this is not within the objectives of this study. These portraits were not intended to provide a detailed analysis of the teaching and learning and it is acknowledged that, mainly due to their broad coverage of a range of issues, they cannot be used to assign causality between specific teaching and learning characteristics and the test results.

On the whole, reference to individual observation records is not included, due both to lack of space in the thesis and the fact that each aspect of the descriptions is derived from several sources. The only exception to this are the discussions regarding 'focus on the forms targeted by the experimental intervention', as it was important to provide more detailed information about how often the teaching outside the experimental conditions focussed on these target forms and the specific activities which this involved.

Descriptions of language covered in the lessons usually adopt the terminology used by the teachers themselves, both to me and the learners, or by the text book or worksheets (with a few exceptions such as 'inflection', 'syntactic'). This is so that the portraits give a clearer picture of the pedagogical and theoretical framework in which the teachers operated, consciously or unconsciously.

4.1.1 General school characteristics

Classes A and C were parallel 'top ability' year 9 classes in an 11-16 Language College. The school is re-applying for this status. All pupils were obliged to study for GCSE French the year following this study, and this may have impacted on their motivation, though it is beyond the scope of this thesis to interpret this further. Departmental meetings were regular and peer observation was carried out, at least once a year for all teachers, by the two senior teachers in the department (one of whom was teacher A). I was aware that there was considerable communication between the teachers in the MFL department regarding lesson plans and resources. There were across year tests in years 7, 8 and 9. The school employed a French assistant and there was an MFL computer suite.

² As defined in chapter 1, this refers to any implicit or explicit, reactive or proactive, attempt to focus learners' attention on specific language items, including metalinguistic description and error correction

Class B was a 'top ability' year 9 class in an 11-18 girls' school. Pupils were not obliged to take GCSE French the following year - seven of the twenty-seven learners in this study chose to continue to study French in year 10, and this may have impacted on the motivation of the learners. Departmental meetings were fairly regular. I was not aware of peer observations occurring at any time within the department.

4.1.2 Class A

Teacher A participated in a collaborative action research project in the year prior to this study (researching the teaching and learning of knowledge and use of German verbs to year 7 pupils). Reports on the action research offer insights into her approach to grammar pedagogy³ and they provide references to observations carried out by others (university, Ofsted and school-based colleagues) which constitute additional testimony to many of the characteristics described below. I was aware that the head of department viewed teacher A as a very reliable and effective teacher.

a) Prior to and during the pre tests

As mentioned in section 3.2.4, most of the learners in class A (and class C) participated in a collaborative action research project on grammar pedagogy for five months in the year before this study, though with another teacher. I was involved in all stages of this project as a research assistant. The learning experience they had is fully documented in the teachers' project report³. The linguistic focus was present and perfect tense verb inflections, though a broader range of inflections was included in the teaching and test materials than in this study. A wide range of FonF/S⁴ techniques were used (for example, see appendices 1a & b). There was a great deal of paradigmatic output practice in short pairwork and teacher fronted Q&A pattern practice, as well as a frequent reactive focus on forms via a variety of error correction strategies. Some of the listening activities in the pre and post test resembled PI materials in that they required learners to attend to the inflection in order to indicate the referent or the tense. This style of activity was also practised at least once. There was no attempt to disguise the linguistic objectives of the tests.

strategies.

³ To maintain anonymity, these cannot be referenced. They are available if requested.

⁴ As defined in section 1.3.1, a mix of Focus on Form and Forms techniques.

The following descriptions refer to observations carried out for the current study.

i) General classroom management

The majority of lessons began within a couple of minutes of the official starting time. In general pupils would have out on their desk the necessary equipment, and would be seated chatting quietly to their neighbours. During whole class work the teacher stood at the front in the middle of the class and, with her eye and head movements, addressed all learners in the room. She usually spoke very quietly in fairly low tones and the atmosphere was generally very calm. She moved up and down the central aisle several times during most lessons and would circulate amongst individuals during individual, pair or group work. Routines were usually strictly adhered to e.g. the pupils were required to sit in the same place every lesson, the four learners who came from a 'middle ability' year 8 class were each seated between pupils who came from the 'top set' year 8 group, the teachers' questions usually required learners to volunteer answers, pupils would raise their hand and wait until the teacher had solicited their response. On most occasions, the teacher repeated the learners' correct response. Meanwhile the other pupils remained quiet and, apparently, attentive. Individual, pair and group work was carried out several times during the observations. After pair or group work, the whole class would go over the work, usually by listening to the work of a few volunteers. There was little fidgeting or off-task behaviour whilst the teacher was talking, though a little more from a few pupils during pair or group work. In general there was tight control over discipline during the lessons. The teacher carried out a few very minor discipline procedures.

Completed homework was almost always worked over together in class and sometimes it was handed in to the teacher and returned marked either the next or the following lesson. The teacher sometimes made comments to individuals during the lesson regarding missing work, giving the impression that she was monitoring the completeness of work in pupils' exercise books.

The teacher gave the impression of having very clear targets about what learners should be achieving in each lesson and also over a longer period of time, in terms of pages to be covered in the text book (for vocabulary and phrase learning) and in terms

of grammar structures. The mapping of these aims onto activities completed in class and at home was clear i.e. the broader targets were realised through specific activities. She designed the scheme of work for all learners of French in year 9 and it was possible to match the work done in class with these schemes. In almost every lesson the teacher wrote "Aims for today" on the blackboard and would sometimes direct pupils' attention to these.

ii) Target language

Pupils were expected to communicate in French with each other and with the teacher during the learning activities where practising the language itself was the objective, and as noted above, time during the lessons was usually on task. There was very little use of French for classroom management and, in terms of IP theory, comprehension of verb inflections was never necessary for learners to understand instructions.

iii) Error correction

An incorrect oral response in class was usually corrected by the teacher using a range of strategies. She would often engage in a metalinguistic discussion in English regarding the error, and if the pupil in question could not resolve the problem, other pupils were asked to contribute. Teacher A's planning was such that errors were frequently pre-empted by phrases such as 'now what do you have to remember here?' 'remember what we were saying last lesson about...'. Models of the expected target-like language were nearly always given to the learners before they would be given the opportunity to use it (perhaps by substituting lexical items). These measures meant that error-correction and metalinguistic discussion were frequently proactive.

iv) Resources & activity types

Many of the activities were from the text book McNab (1994). The activities were sometimes adapted a little and supplemented with notes on the board and/or whole class oral practice using flash cards or model dialogues for rehearsal. Worksheets were given to the pupils only occasionally. Usually activities used from the text book emphasised the function of fixed phrases. Listening and reading activities required learners to focus on lexical items. Speaking and production activities generally required accuracy. Most lessons involved some, fairly limited, reactive metalinguistic discussion/explanation, led by the teacher. However, several lessons

had a proactive focus on form objective and these usually involved a small number of forms in the activities and discussions. Occasionally more than one whole paradigm was covered. Approximately once every fortnight the Foreign Language Assistant (FLA) took small groups of learners out of the lesson, usually in order to rehearse planned Q&A routines. The learners used the IT suite twice before the pre tests - both times for word processing some work they had already planned or handwritten.

v) *Focus on the forms targeted by the experimental intervention*

A few of the lessons prior to the pretests involved a considerable amount of focus on the 'forms targeted by the experimental intervention' (henceforth, 'target forms' or 'target features') - 17 Sept, 5, 10 & 24 Oct. The first two of these involved learners' production of present tense verb inflections and the second two the production of perfect tense constructions. However, in between and during each of these lessons there were other objectives e.g. one of the lessons focussing on the perfect tense (1st & 3rd person singular, avoir, être and irregular verbs) also involved practising a range of fixed phrases in order to arrange to go out (e.g. on pourrait, est-ce que...), following the structure suggested in McNab (1994) pp16-26. On 24 Oct the FLA took out groups of about 6 learners to practise perfect tense verbs. A dice was used where shaking a 1 elicited first person singular inflections, a 2 2nd person singular etc.

When six (of the eight) narrative level oral pre tests were being carried out with learners from this class, the ongoing lesson from which learners came involved explicit grammar presentation and revision practice on the perfect tense, though further details are not known. It was noted that some learners seemed to pause before or during their productions for longer than in subsequent tests and more than learners in other classes. However, no systematic measurement was taken of this. It is pointed out that any effect this may have had on pretest scores is likely to have affected both PI and EI learners alike.

b) *During the intervention*

All the PI and EI activities were completed in every session, except on one occasion when the last EI activity out of 6 was not done. I believe that teacher A read through all the intervention materials prior to using them (e.g. she made comments prior to the activities). Teacher A adhered very closely to the guidelines I provided for carrying

out the PI and EI materials. The arrangements for the intervention (e.g. for rooms and equipment) were carried out smoothly every lesson.

For administrative reasons, there were two lessons and four parts of lessons where the teaching was not part of the experimental intervention. From observation records, teachers' comments and lesson plan notes, it is known that these lessons did not involve any proactive FonF/S concerning the present and perfect tenses.

c) *After the intervention - between post and delayed post tests*

The general classroom management style described above was maintained between the post and dp test.

The programme of work centred on the teaching of the imperfect tense and contrasting its use with the perfect tense. There were many examples of metalinguistic discussions around the formation of the imperfect inflections in a wide range of regular (er, ir, and re) and irregular verbs. Many of the activities involved some kind of written translation, and, for interpreting the French, there was almost always a possibility for learners to use some cue other than the intended verb inflections (e.g. by using 'pendant que' as a lexical cue or by noticing the lexical semantics of the verb). Two lessons (with no observation records but according to self report from the teacher and the worksheets used) consisted entirely of explicit FonFs. The first lesson provided a description of the formation of the 3 regular present tense paradigms and the perfect tense including the 3 regular past participle formations and both avoir and être verbs. The learners also had about 10 minutes to translate 16 sentences into French, 8 labelled 'le présent' and 8 labelled 'le passé composé'. The teacher then read out the correct French translations. The second lesson consisted of a similar approach for the imperfect tense and ways of expressing the future without the future tense. Throughout the lessons between post and dp test there was a heavy emphasis on written production. On the rare occasions where listening and reading work were carried out, learners' attention was directed to lexical items.

4.1.3 *Class B*

Almost all of the learners in class B were also taught by teacher B the year before this study. However, these descriptions are based solely on observations carried out for the current study. I was aware that the head of department was pleased with the objectives that teacher B set, the enthusiasm she brought to the planning process, the energy she conveyed in lessons and her apparent popularity with pupils.

- a) *Prior to and during the pre tests*
- i) *General classroom management*

The majority of lessons began at least 10 minutes after the official starting time, during which time pupils arrived, chatted to their friends and stood around the classroom. During whole class work teacher B stood at the front to the left/middle of the room. She almost always spoke loudly or very loudly and the atmosphere was usually very animated, the teacher sometimes banging the table in front of her to get pupils' attention. Occasionally, teacher B moved along the rows of pupils, paying attention to individuals and small groups. Some pupils were more difficult to reach as there was no central aisle and a few pupils on the back row would not often interact with the teacher. Individual, pair and group work was carried out several times during the observations, with considerable on-task interaction between the pupils.

Classroom routines were not used systematically e.g. generally the pupils sat where they liked; homework was given frequently but often explained after the lesson bell had rung and many pupils' attention was elsewhere; I was not aware of systematic procedures to pursue incomplete class and home work⁵; although the queries which pupils shouted out were often not relevant to the task in hand, the teacher almost always responded to the query. During Q & A sessions, most often pupils would call out an answer whilst or without raising a hand. Teacher B often answered her own questions. While the teacher talked to the whole class, there was frequently considerable chat and fidgeting amongst groups of pupils. There was often apparent confusion amongst the pupils over criteria to be used for marking their own work.

⁵ though a frequent general instruction in lessons was to do/redo/complete homework

Teacher B had broad, communicative / functional / task-based objectives, based on planning and carrying out a visit to France which, the pupils were told, would happen at the end of year 9. She expressed the aims of each lesson to the learners in terms of which worksheets or specific tasks (e.g. a letter to the tourist information office) had to be finished. This sometimes implicitly involved a particular set of vocabulary or phrases. However, teacher B's objectives rarely manifested themselves in specific sets of morphosyntax to be learnt via a thread of activities.

ii) Target language

Pupils were expected to use French at moments where practising the language itself was the objective. Some classroom management was carried out in the target language and there were several incidents where the teacher used French in order to respond to pupil queries. However, the French used was not usually essential to the actual message and, in terms of IP theory, comprehension of verb inflections was never necessary for learners to understand classroom events.

iii) Resources & activity types

The teacher very frequently gave out worksheets which were taken from a range of sources. Text books were never used in the lessons. Most pupils had a vocabulary book, and some pupils had written a few words in it. On two occasions the class worked in the IT suite, searching for tourist information via a website they were directed to. There was a strong emphasis on written work and on written and spoken accuracy. Reading and listening activities usually required learners' attention to be focussed on lexical items and phrases with fixed transactional functions. Teacher B carried out one gap fill reading activity which she said was to train the learners to realise that knowing the "grammatical category" of word/s was necessary for such activities. However, the activity could actually be done by relying on semantic cues and so the assignation of grammatical categories became a general cognitive activity, with little or no relation to meaning. Teacher B frequently referred to mnemonics intended to remind pupils of a grammatical 'rule'. She also relied heavily on making notes on the board, often covering it with notes, diagrams and phrases.

iv) *Error correction*

Error correction was frequent, both in written and oral work. The teacher would tend to provide an explicit commentary in English, including metalinguistic explanations, translation, transliteration and mnemonics, when learners produced non-target-like language. Learners were usually expected to produce one or two words, rarely full sentences, in front of the class. The teacher would usually repeat correct responses. Incorrect responses were usually corrected in some way, often by repeating the question until a learner called out the right answer or the teacher provided a recast.

v) *Focus on the target forms*

The observation records suggest that there was very little, if any, teaching of the target features prior to the pre tests. The lessons were mainly driven by topic and/or functional language and at the start of most lessons teacher B began with fairly communicative goals. However, a wide range of grammatical features were discussed in lessons. Some of these appeared to be incorporated proactively and were presented didactically: prepositions (en face de, à côté de), isolated forms of particles with determiners (du, au, de l', d'un), question formation and modal verbs + infinitives. The vast majority of explicit grammar discussions arose reactively. Teacher B was frequently distracted from initial goals into quite complex form/function explanations in English and from checking comprehension of one grammatical explanation to providing another.

There was fairly extensive usage by the teacher of metalinguistic terminology, including '2nd clause', 'subject', 'conditional', 'article', 'negation', 'imperfect'. Pupil use of these terms was usually inaccurate, or appeared to be a 'trial and error' listing of terms. Sometimes pupils asked each other or the teacher what they meant after the terms had been used.

b) *During the intervention*

All the PI and EI activities were carried out, except that two EI activities (in different units) were not completed by all pupils. In the first two units, teacher B imposed her teaching style a little on the materials i.e. expanded the EGI and provided fuller explanations when checking the responses at the start of the PI referential activities. She then adhered more closely to the guidelines.

Class B had one entire lesson and two slots of about 25 minutes with no experimental intervention. In these lessons, the activities were carried out with a non-specialist teacher (as teacher B was absent) and they did not involve any FonF/S on the target features.

c) After the intervention - between post and delayed post tests

The characteristics of the lessons remained similar to those described prior to the pre test.

The lessons were topic and task driven, with chunks of language and their English equivalents being presented to and then used by the learners in order to perform certain functions related to eating in a restaurant. First learners worked on lexical items on menus in reading and writing tasks, and then they wrote a text message to a French friend inviting them to a restaurant. The major task during this period was to prepare (i.e. write down and rehearse during several lessons) a role play in a café, following a model with key phrases provided (including the verb). The role plays were then performed over two lessons, one of which I observed. The learners either read their script or had memorised it. Finally, they wrote a letter to a friend describing what the meal was like (following a template letter). The two last activities involved the teacher providing an explanation of the formation and use of the imperfect tense and some learners used a couple of fixed phrases in the imperfect in their written and oral work. Towards the end of this period the pupils worked individually or in small groups through several GCSE role plays (i.e. prepared their responses in writing) and two GCSE foundation reading papers. In addition, some worksheets (reading and writing tasks) on the topic of 'buying presents' were completed.

During this time, the observation notes and comments from teacher B show that specific teaching of the present or perfect tenses did not feature in the lessons.

4.1.4 Class C

As described in section 3.2.2, class C had three teachers during the course of this study: for 3 lessons per fortnight they had teacher C then teacher D (after the post

this portrait refers to *the teacher who took the majority of lessons* i.e. teacher C then teacher D.

a) Prior to and during the pre tests

As described in sections 3.2.4 and 4.1.2, almost all of the learners in class C had, in the previous year, received extensive FonF/S instruction regarding present and perfect tense verb inflections over a period of five months (as received by a parallel class that was participating in an action research project).

The following descriptions refer to observations carried out for the current study.

i) General classroom management

The majority of lessons began about five minutes after the official starting time, though this varied considerably. As pupils entered they would generally get out the necessary equipment and chat quietly to their friends. During whole class work the teacher stood at the front in the middle of the class and would usually move up and down the two aisles to monitor individual and group work. The atmosphere was generally quite calm though it could be very lively at times. Routines were fairly strictly adhered to e.g. the pupils sat in the same place every lesson; there was a register routine in most lessons where learners had to provide a phrase linked with the current topic when their name was called out (e.g. on peut + infinitive); the teachers' questions usually required learners to volunteer answers (occasionally she would name individuals who had not volunteered) and pupils would generally raise their hand and wait until the teacher had solicited their response. Pupils were usually required to say one or two words in response to the teachers' questions, though occasionally they had to use a whole phrase, particularly when carrying out the register routine described above. There was a little fidgeting and off-task chat whilst the teacher or other pupils were talking. Individual, pair and group activities were carried out several times during the observations and pupils generally worked conscientiously. After pair or group work, the whole class would go over the work, usually by listening to the work of a few volunteers.

The teacher gave the impression of having clear targets each lesson in terms of activities and pages to be covered in the text book for vocabulary and phrase learning.

Occasionally teacher C had more grammatical targets (e.g. paradigms of tenses).

Usually the teacher wrote "Aims for today" on the blackboard and she would sometimes direct pupils' attention to them.

Homework was usually set from the text book e.g. true / false exercises, matching beginnings and endings of sentences, and was usually pursued in some way.

ii) Target language

The target language was more prominent in the teacher's management talk than in the other two classes. Pupils were expected to communicate with each other in French during activities where practising the language itself was the objective and also to ask permission to take off their blazer, and occasionally for one or two other classroom requests. The majority of teacher E's management talk was in English. With both teachers learners did not often have to understand the French in order to respond or carry out the activity (other cues could be used such as context, paralinguistic features, or peer or teacher translation or modelling). A typical feature of both teachers' talk was a matrix of English sentences with a few key nouns in French, particularly cognates. In terms of IP theory, it was rarely if ever essential to understand the meaning of verb inflections in order to understand the message.

iii) Error correction

Oral error correction was not a very prominent feature of the lessons with teacher C, though more so with teacher E. Teacher C rarely gave an explicit correction and explanation of the error, though this was more common with teacher E. With both teachers, most oral and written production was based around fixed phrases with 'replace the slot' type responses, and this probably reduced the amount of non-target like language produced.

iv) Resources & activity types

The teaching was generally driven by the topics and some of the grammatical structures suggested in the text book McNab (1994). One of the main topics was 'the environment', including descriptions of seasons, global environmental problems and what the pupils did or should do to protect their environment. The activities were sometimes adapted a little and supplemented with notes on the board or worksheets.

Usually the materials presented fixed phrases and provided activities which emphasised their function. Speaking and production activities generally required accuracy in terms of verb inflections and were based on models given to the learners e.g. drawing a poster about environmental protection with an imperative instruction, copied from a range of phrases provided. Production activities occupied most of class time and homework activities. Listening and reading activities required learners to demonstrate their understanding of lexical items. There was relatively little metalinguistic discussion/explanation. Approximately once every fortnight the Foreign Language Assistant (FLA) took small groups of learners out of the lesson, usually in order to rehearse planned Q&A routines.

v) *Focus on the target forms*

There was some proactive and reactive FonF/S teaching, some of which involved the target features in this study. Forms that received proactive focus were present imperative vous form of verb, verb (peut, il faut) + infinitive, je + present verb, and a wide range of perfect tense inflections. Two lessons in particular involved a considerable amount of proactive focus on the target forms - 7 December (not observed) and 12 December (observed). On 7 December, teacher C left some cover work involving worksheets which required the written production of the perfect tense. On 12 December, the first 15 minutes aimed to generate conceptual thinking regarding the uses of avoir and être, by asking pupils what 'ideas/things' could be expressed using avoir and être e.g. nationality, personality. The learners then had to underline avoir and être verbs in a passage written in the present, before carrying out a dictation about a weekend routine in the present tense. They were then asked to transform this into the perfect tense, which the whole class carried out together, some pupils referring back to notes made the previous year. The teacher briefly reminded the learners of the mnemonic 'MRVANSTRAMPED' for deciding when to use être in the perfect tense. The writing was completed for homework. Teacher C told me that this work was to prepare for some writing, the following lesson, in the perfect tense using both avoir and être as auxiliaries, which would be assessed using National Curriculum criteria.

b) During the intervention

The description above remains relevant for the lessons observed during the intervention, in terms of classroom management, use of target language and activity type. The lessons continued to be driven mainly by the topics and activities presented in the text book. The topics covered were: parts of the body, illness, teenage problems, food likes & dislikes, healthy eating, sports, describing personalities, at the campsite and describing one's school bag. As can be seen from McNab (1994) pp 54-84, a wide range of vocabulary and functional phrases was involved during these activities (and one lesson contained some explicit teaching and mechanical written practice on the use of particles with determiners e.g. du, des). However, there was no focus on the target forms between the pre and post tests.

c) After the intervention - between post and delayed post tests

The general classroom management (e.g. time on task) remained broadly similar to the description above for teacher C. One slight difference was that teacher D wrote the aims for the lesson on the board more frequently than teacher C. Also, teacher D did not continue the 'register routine' where pupils would practise a structure / phrase after their names had been called. Teacher D engaged in a few more 'naturalistic' question and answer sessions with the pupils in the target language, where learners were not told in advance the meaning of the teachers' questions and were not given a model answer. Teacher D used more target language than the other teachers, for instructions and whole class question and answer sessions.

Error correction was a more prominent feature of teacher D's lessons than teacher C's. Teacher D was more willing to let pupils try to produce language for which they have no model, which elicited more non-target-like language than in teacher A and C's lessons. Teacher D used a range of error correction strategies, for example recasting or, more implicitly, repeating the error with raised intonation. She did not often insist that the learner articulate the correct form.

The text book continued to lead the planning of vocabulary and grammatical structures. Topics covered this term included freetime, holidays, clothes and prices (including a 'fashion show' where pupils brought in different clothes and their friends described what they were wearing).

Teacher E had three lessons with class C between post and dp test. One included a grammar practice worksheet filing in blanks with 'du', 'des', 'de la'; another a narrative in the perfect tense describing some holidays where pupils had to correct person and number errors in the verbs that were underlined; and the third, in the IT suite, where learners had to find information regarding the profiles of various football stars from the internet, coinciding with the World Cup.

With all teachers reading and listening tasks always focussed learners' attention on lexical items (including during the FonF/S activities described below). There were very few, if any, moments in the lessons where the tasks were such that (intentionally or not) learners *had* to attend to the target features of form in order to complete the task.

Between the post and dp test there was a considerable amount of FonF/S. This included the syntax and agreement of adjectives and future expressions using 'aller'. Most relevant to this study is that a considerable amount of class and homework time was given to learning the perfect and, to a lesser extent, imperfect tenses. The emphasis was on written production, though some oral production, of 'avoir' and 'être' regular and irregular verbs in the perfect tense. In one reading activity learners had to underline verbs that were in the perfect tense (for this, they could look for specific forms without attaching meaning to them). Another reading activity asked learners to pick out, from a written narrative, the French for some English past tense phrases (and this could be done by finding the relevant verb stem or temporal adverb). There were no taped listening activities. Teachers D and E and the FLA said a few sentences to the learners in the perfect tense. The reading and listening activities did not resemble PI activities: the interpretation of the verb inflections, in terms of the tense, person or number, was rarely essential to the task. The exceptions were when the teacher directed learners' attention to a written verb (ending in 'ez') and asked "am I talking to one of you or all of you?" and, on another occasion, the teacher said "[the FLA] va chanter - what tense?" and "nous allons écouter - what tense?". These bear some resemblance to referential PI activities, as the learners probably listened to verb inflections in order to get the required information, although it is possible that they were able to guess from contextual clues.

4.1.5 Summary of 4.1: Portraits of the lessons

One of the aspects of the lesson portraits most relevant to this study was whether and how the target features in this study occurred in the lessons. The following is a summary of this for each class⁶.

Prior to the pre tests

Class A: considerable FonF/S (extensive in the previous year)

Class B: no FonF/S (some in the previous year)

Class C: some FonF/S (extensive in the previous year)

Between the pre and post tests

Classes A and B: no FonF/S outside the experimental protocol

Class C: negligible FonF/S

Between post and dp tests

Classes A and C: extensive FonF/S (for perfect tense)

Class B: no FonF/S

This suggests that the experimental design had reasonable internal validity in that in classes A, B and C any differences found between the pre and post test are likely to be due to the experimental protocol (i.e. PI and EI in classes A and B and no FonF/S in the non-active control class C).

Between the post and dp tests, in classes A and C, any learning gains (or apparent maintenance of gains made between pre and post tests), may be due to the FonF/S received in lessons with their class teachers. In class B any gains made, or maintained from post test, are unlikely to be due to lessons after the post test, as there was no FonF/S once the experimental intervention had finished. They are therefore most likely due to the experimental intervention (PI and/or EI).

The portraits also implied that classes A and C may have better metalinguistic knowledge of the target structures than learners in class B, although this is a fairly impressionistic judgement based on observations of the lessons prior to the pre test and data from the action research project.

⁶ In this summary, FonF/S refers to the target forms of *this study*

The lesson observations indicated that the learners had not been exposed to instruction resembling PI before this study. It was suggested that listening and reading activities rarely, if ever, required learners to interpret the meaning of language items of low communicative value such as verb inflections.

These portraits have suggested that learners in class B are less accustomed to tightly planned sequences of activities based around a small set of grammatical features. Class B learners are more accustomed to lessons based around functional tasks (e.g. writing a letter to a tourist office), where most key phrases are provided for them, and a wide range of language features may be discussed reactively at any point in the lesson. In contrast, learners in class A and (perhaps to a lesser extent) class C are accustomed to class teaching and research projects in which a small set of grammatical features are the explicit focus of an intensive sequence of activities. It is considered that these learners are more likely than learners in class B to expect activities to be structured so that a few language features are made salient (in some proactive way) with the aim of the pupils attending to them and being asked / tested about them by the teacher or subsequent activities/tests. Henceforth, this is referred to as an 'ethos of teaching and testing grammar'.

4.2 Assessing the similarity of the classes at the outset

It was important to see whether the different classes were comparable at the outset in terms of the pre test scores in all measures: listening, reading, writing and speaking (see sections 3.6 and 3.8 for descriptions). Oneway ANOVAs and post hoc tests were carried out on the pre test data in order to ascertain the statistical significance of any initial differences between the classes⁷. The results of the ANOVA and post hoc tests are presented in tables 4.1 and 4.2 respectively. It is acknowledged that two of the conditions required for these tests were not met: the variances of the different classes and groups were not equal except for the speaking pre test scores (according to Levene's test, see section 3.8.5 and appendix 32) and some of the sample distributions could not be considered normal (see chapters 5 and 6 and appendix 29). Non-parametric tests were therefore also carried out (Kruskall Wallis and Mann-Whitney

⁷ See section 3.8.5 for a description of these tests. The post hoc tests used the Games Howell adjustment for multiple comparisons as the variances and sample sizes were unequal, except for the speaking test scores where the Sidak adjustment was used as variances were equal and this correction has good control over the Type I error rate (Field 2000 p276).

U tests, see section 3.8.6 for a description). However, as they produced similar findings to the parametric tests, only the parametric tests are presented here⁸. Non-parametric tests were carried out for the speaking test results for the additional reason that the n was small. One aspect of this analysis differed from the parametric test results and is presented below.

Pre test measure		Sum of Squares	df	Mean Square	F	Sig.
Listening	Between Groups	1804.962	2	902.481	5.000	.009
	Within Groups	14801.808	82	180.510		
	Total	16606.769	84			
Reading	Between Groups	11494.584	2	5747.292	24.242	.000
	Within Groups	19440.364	82	237.078		
	Total	30934.948	84			
Writing	Between Groups	8641.937	2	4320.969	13.444	.000
	Within Groups	26355.066	82	321.403		
	Total	34997.003	84			
Speaking	Between Groups	764.400	2	382.200	2.106	.133
	Within Groups	8346.332	46	181.442		
	Total	9110.732	48			

Table 4.1 One way ANOVAs to compare the pre test scores in all measures in different classes

Pre test measure	(I) CLASS	(J) CLASS	Mean Difference	Std. Error	Sig.	95% Confidence Interval	
			(I-J)			Lower Bound	Upper Bound
Listening	A	B	10.7420	3.70609	.016	1.7535	19.7304
	A	C	1.9155	3.92302	.877	-7.5603	11.3912
	C	B	8.8265	2.99242	.013	1.6151	16.0379
Reading	A	B	24.7189	3.63686	.000	15.8451	33.5927
	C	A	.4937	4.68694	.994	-10.7907	11.7782
	C	B	25.2127	3.63897	.000	16.3575	34.0678
Writing	A	B	14.6718	3.54402	.001	6.0220	23.3216
	A	C	-10.1864	5.44245	.157	-23.3068	2.9340
	B	C	-24.8582	4.55761	.000	-36.0178	-13.6985
Speaking	A	B	5.0915	4.69182	.632	-6.5339	16.7169
	A	C	-4.5262	4.76238	.721	-16.3264	7.2740
	B	C	-9.6177	4.69182	.132	-21.2431	2.0077

Table 4.2 Post hoc tests following one way ANOVAs to compare pre test scores in all measures in different classes

⁸ See appendix 27 for the non-parametric tests

It was found that, for all measures except the speaking, the differences at pre test found by the ANOVA were due to differences between classes A and B and between classes C and B: class B started from a statistically significantly lower baseline score than the other two classes and classes A and C had the same (statistically speaking) pre test scores.

Although the ANOVA found no statistically significant differences between the classes at pre test in the speaking scores, post hoc tests were carried out to explore the data for trends. A one-tailed test of the hypothesis that class C's scores were statistically significantly higher than class B's found that the difference was approaching significance at the 95% confidence level ($p=0.66$). In support of this, the Mann Whitney U test found that class C had a statistically significantly higher speaking pre test score than class B, (81.000, Wilcoxon W 234.000, Z -1.982, $p=0.025$). It is noted that this is the only measure where there are no statistically significant differences between classes A and B, and where a two-tailed test is not sensitive enough to register a statistically significant difference between classes B and C. These findings for the oral production measure are perhaps indicative of the fact that verbal inflection in oral performance is usually late to develop compared to other measures.

This analysis has suggested that classes A and C were from the same population at the outset of the study, reinforcing the fact that class C's scores are a useful comparison for class A's scores, at least between pre and post test where there was negligible FonF/S on the target features in class C.

This analysis has also shown that class B was probably from a different population to classes A and C at the outset in terms of aural and written interpretation of and written production of verb inflections, despite being from a similar general educational stage and establishment. This suggests that this study will provide some understanding of possibly differential impacts of PI and EI at different developmental stages. It has also suggested that to ensure that initial differences between class B and the other two classes did not affect the analysis of all the data together, additional analyses should be carried out for each measure (see chapters 5 and 6 and appendices

33 and 34)⁹. First, *separate* analyses were conducted on the two intervention classes (A and B). Second, analyses were carried out on classes A and B together but without class C, as it was not directly necessary for the comparison of PI with EI (and, as shown in section 3.3, it was not comparable with the PI and EI merged groups at the outset). Third, analyses were conducted on the *gains* made by different groups and classes of learners between each pair of tests (in addition to analyses of the actual scores).

4.3 Analysing the 'attitude data'

An analysis of the attitude data was necessary to explore whether the learners and teachers preferred one instructional approach (PI, EI or class C) to another. This will contribute to our understanding of whether the pattern of results in the achievement tests were correlated with any pattern of reported preferences, dependent on, for example, class or instructional groupings. Such a correlation could imply that 'motivation' (pupils' and or teachers) may have contributed to any influence PI and EI had on the learners' test scores. Three sets of data are used in this section: Likert scales administered after each set of tests, questionnaires administered to PI and EI learners just before the post tests, and semi-structured interviews with the teachers following the post tests (see chapter 3 for details of the collection of this data).

4.3.1 Likert scales data

The two short Likert scale questions were used to give a very broad indication of the attitudes of the learners towards French and whether these changed to a major extent during the intervention and between the post and delayed post test. It is acknowledged that many factors could cause a change in expressed preferences of a subject and perception of difficulty. For example, in class C, there was a change of teacher after the post test; in school 2 French was not an obligatory subject in year 10. The purpose of the data elicitation was to confirm that there had been no marked change which correlated with experimental instructional type (EI or PI).

⁹ As the experimental groups were equivalent at the outset (section 3.3 and appendix 15 show how the randomised matched pairs sampling produced homogenous groups) an analysis of all three classes together was valid. It was necessary to include class as a factor to explore its influence on the impact of PI and EI in different classes.

One set of Likert scales (appendix 25) asked learners to rate their preference for French compared to other subjects according to 5 phrases with the first saying French was the favourite and the fifth saying it was the least favourite. The other set asked them to rate the difficulty of French compared to other subjects, with the first phrase saying French was the most difficult and the fifth phrase saying they found it the least difficult. The results are presented in tables 4.3 and 4.4.

Data from Likert scales can be analysed as categorical, rank-ordered, or continuous scale data (Brown 1988, Hatch & Lazaraton 1991). These would entail, respectively, chi square, another non-parametric test (such as Friedman or Wilcoxon signed ranks test) or parametric tests (such as t tests or ANOVAs). The Friedman and Wilcoxon tests were most suitable for analysing this data¹⁰. Both these tests are 'assumption free' in that they do not require normally distributed data and can be used with ranked data (see 3.8.6 for a description of these tests).

¹⁰ Chi square tests require that each subject only contributes to one cell (i.e. not repeated measures) and that the expected frequency in all cells is above 5 (see Field 2000 p66-9 and Jones & Bryant 1998 p62). However, this data was repeated measures (each subject had 3 measures) and, even with merged categories (e.g. making one category out of learners who selected either phrase 1 or 2), some of the cells in some of the analyses fell below 5. The t tests and ANOVAs require interval data and for it to be normally distributed - neither of these conditions were met by this data¹⁰.

a) Learners' reported preferences for learning French compared to other subjects

Test	Grouping	1 (favourite)	2	3	4	5 (least favourite)	total	Mean	Std. Dev.
Pre	Overall	1	16	38	21	6	82	3.18	0.88
Post	Overall	0	23	34	21	8	86	3.16	0.93
Dp	Overall	0	14	27	25	10	76	3.41	0.94
Pre	Class A	0	7	14	5	0	26	2.92	0.69
	Class B	0	3	9	9	6	27	3.67	0.96
	Class C	1	6	15	7	0	29	2.97	0.78
Post	Class A	0	12	12	5	0	29	2.76	0.74
	Class B	0	5	6	8	8	27	3.70	1.10
	Class C	0	6	16	8	0	30	3.07	0.69
Dp	Class A	0	8	12	5	0	25	2.88	0.73
	Class B	0	4	3	8	10	25	3.96	1.10
	Class C	0	2	12	12	0	26	3.38	0.64
Pre	Merged EI	0	5	7	9	4	25	3.48	1.00
	Merged PI	0	5	16	5	2	28	3.14	0.80
Post	Merged EI	0	9	6	6	6	27	3.33	1.18
	Merged PI	0	8	12	7	2	29	3.10	0.90
Dp	Merged EI	0	5	6	4	7	22	3.59	1.18
	Merged PI	0	7	9	9	3	28	3.29	0.98
Pre	Class A EI	0	4	4	4	0	12	3.00	0.85
	Class A PI	0	3	10	1	0	14	2.86	0.53
Post	Class A EI	0	7	4	3	0	14	2.71	0.83
	Class A PI	0	5	8	2	0	15	2.80	0.68
Dp	Class A EI	0	3	6	2	0	11	2.91	0.70
	Class A PI	0	5	6	3	0	14	2.86	0.77
Pre	Class B EI	0	1	3	5	4	13	3.92	0.95
	Class B PI	0	2	6	4	2	14	3.43	0.94
Post	Class B EI	0	2	2	3	6	13	4.00	1.15
	Class B PI	0	3	4	5	2	14	3.43	1.02
Dp	Class B EI	0	2	0	2	7	11	4.27	1.19
	Class B PI	0	2	3	6	3	14	3.71	0.99

Table 4.3 Results of the Likert scales regarding pupils' reported preference for learning French

Friedman tests (Table 4.4) were carried out to find if there were statistically significant differences between different sample groupings' expressed preferences over the three tests.

	Overall	Class A	Class B	Class C	Merged EI	Merged PI	Class A, EI	Class A, PI	Class B, EI	Class B, PI
N	72	22	25	25	20	27	9	13	11	14
Chi-Square ¹¹	6.748	.933	4.000	8.150	3.500	.667	4.800	.200	2.571	1.636
df	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	.034	.627	.135	.017	.174	.717	.091	.905	.276	.441

Table 4.4 Friedman tests to find differences in learners' reported preference for learning French

The results in table 4.4 show no statistically significant impact of the type of instruction (PI or EI) on the pupils' reports of how they liked French compared to other subjects, regardless of which class learners were in (A or B).

In contrast, in the overall dataset and in class C there were significant changes in pupils' preference ratings. Paired Wilcoxon signed ranks tests were carried out to find the source of these differences (Table 4.5).

Sample group		pre - post	post - dp	pre - dp
Overall scores	Z	-.308	-2.874	-2.141
	Asymp. Sig.	.758	.004	.032
Class C	Z	-1.414	-2.309	-2.352
	Asymp. Sig.	.157	.021	.019

Table 4.5 Two-tailed Wilcoxon tests to compare differences between pairs of tests in pupils' reported preference for learning French

It was found that the difference in the overall preference scores was due to changes between the post and dp tests and between pre and dp tests. The direction of this difference was that there was a decrease in the learners reporting that they liked French. There were no differences between the pre and post tests, reinforcing the finding above that the experimental intervention (PI and EI) did not have a marked impact on learners' reported attitude to French lessons.

For class C the Wilcoxon tests showed that the differences were also due to changes in expressed preferences between post and dp test and between pre and dp test. It is

¹¹ The Friedman test uses the chi-square statistic, but it does not use Chi-square test procedures.

suggested that this was due to the change of teachers in class C just after the post tests and this was supported by informal pupil reports about their French lessons between post and dp tests. It is suggested that the change in the overall dataset's preference ranking may have been due to these changes in class C.

Although these findings do not have any direct impact on the interpretation of the findings in this study, it is useful in that it has suggested that although these Likert scales are relatively blunt instruments, they did seem to reflect this change of mood. This adds strength to the finding from these analyses that between pre and post tests there were no significant changes in any group of pupils' reported preferences.

b) Learners' reported perception of the difficulty of learning French compared to other subjects

Test	Grouping	1 (most difficult)	2	3	4	5 (least difficult)	Total	Mean	Std. Dev.
Pre	Overall	4	38	31	8	1	82	2.56	0.79
Post	Overall	3	31	45	7	0	86	2.65	0.68
Dp	Overall	2	38	33	3	0	76	2.49	0.62
Pre	Class A	0	8	16	2	0	26	2.77	0.59
	Class B	4	16	4	3	0	27	2.22	0.85
	Class C	0	14	11	3	1	29	2.69	0.81
Post	Class A	0	6	20	3	0	29	2.90	0.56
	Class B	3	13	10	1	0	27	2.33	0.73
	Class C	0	12	15	3	0	30	2.70	0.65
Dp	Class A	1	8	16	0	0	25	2.60	0.58
	Class B	1	14	8	2	0	25	2.44	0.71
	Class C	0	16	9	1	0	26	2.42	0.58
Pre	Merged EI	2	11	10	2	0	25	2.48	0.77
	Merged PI	2	13	10	3	0	28	2.50	0.79
Post	Merged EI	3	7	15	2	0	27	2.59	0.80
	Merged PI	0	12	15	2	0	29	2.66	0.61
Dp	Merged EI	2	10	10	0	0	22	2.36	0.66
	Merged PI	0	12	14	2	0	28	2.64	0.62
Pre	Class A EI	0	3	8	1	0	12	2.83	0.58
	Class A PI	0	5	8	1	0	14	2.71	0.61
Post	Class A EI	0	3	10	1	0	14	2.86	0.53
	Class A PI	0	3	10	2	0	15	2.93	0.59
Dp	Class A EI	1	4	6	0	0	11	2.45	0.69
	Class A PI	0	4	10	0	0	14	2.71	0.47
Pre	Class B EI	2	8	2	1	0	13	2.15	0.80
	Class B PI	2	8	2	2	0	14	2.29	0.91
Post	Class B EI	3	4	5	1	0	13	2.31	0.95
	Class B PI	0	9	5	0	0	14	2.36	0.50
Dp	Class B EI	1	6	4	0	0	11	2.27	0.65
	Class B PI	0	8	4	2	0	14	2.57	0.76

Table 4.6 Results of the Likert scales for learners' reported opinions regarding the difficulty of French

Friedman two-way analysis of variance tests were carried out to assess whether there were any changes in the opinions of learners between the different tests (see Table 4.7).

	Overall	Class A	Class B	Class C	Merged EI	Merged PI	Class A, EI	Class A, PI	Class B, EI	Class B, PI
N	72	22	25	25	20	27	9	13	11	14
Chi-Square	2.641	2.364	5.019	3.000	1.805	1.733	2.000	.857	2.138	4.750
df	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	.267	.307	.081	.223	.406	.420	.368	.651	.343	.093

Table 4.7 Friedman tests to find differences in learners' reported perception of difficulty of French.

This analysis has suggested that over the course of the study, learners' reported perception of the difficulty of French compared to other subjects, according to the Likert scales used, did not change significantly, regardless of class and/or group. This implies that PI and EI did not alter learners' sense of achievement or progression. This, to some extent, suggests that any differences found in the test scores at post and dp test are unlikely to be due to motivational issues. The next section provides more quantitative and qualitative evidence from the questionnaires supporting this claim.

4.3.2 Pupils' questionnaires - classes A & B

The data collected from the questionnaires (appendix 26) was analysed to obtain an impression of whether the two types of activities (PI and EI) were different in terms of enjoyment and sense of learning felt by pupils¹². This was necessary to explore whether they may have motivated learners in different ways.

Although I asked pupils to complete the questionnaire in silence with their own opinions, there were a handful of incidents where learners did communicate with each other, and in the EI group in class B a couple of learners wrote identical comments. Nevertheless, due to the infrequency of this and the impossibility of being able to tell whether the copied comments were genuinely felt, these have been included as separate opinions.

¹² As discussed in section 3.7.2, any pollution of the data caused by the fact that I was involved in teaching and handing out the questionnaires, or by the learners' awareness of their involvement in research, these are likely to have affected both PI and EI learners equally. Section 3.7.2 also discusses measures taken to avoid such pollution.

A tabular summary of the pupils' responses to each question is provided in Table 4.8. Appendix 37 presents further quotes from the pupils' responses, in particular regarding whether they felt PI and EI had been helpful and changed the way they noticed French grammar.

	EI groups		tot	PI groups		tot
Question	Class A (n=14)	Class B (n=13)		Class A (n=15)	Class B (n=13)	
<i>Enjoyed?</i>	10 (4 didn't - 1 due to friendship groups, 2 because boring, 1 'don't like tests')	10 (3 learners didn't - 2 due to friendship groups, 1 said she didn't like French)	20	15 (of these, 3 said also boring)	11 (1 didn't due to friendship groups, 1 because 'boring').	26
<i>Helpful & useful?</i>	14	10 - (3 didn't - 2 identical comments)	24	14 (1 didn't because 'knew it already')	13	27
<i>Different to normal?</i>	13 (1 no)	11 (2 no - identical comments)	24	13 (2 no)	13	26
<i>Should activities continue?</i>	11 (2 no, 1 no response)	9 (4 no - 2: 'not all the time', 2: identical comments)	20	13 (2 no)	11 (2 no)	24
<i>Positive opinions</i> ¹³	12	11	23	13	6	19
<i>Negative opinions</i>	2	6	8	2	4	6
<i>Changed noticing of grammar?</i>	12 (2 didn't)	8 (4 didn't - one was highest pretest score)	20	14 (1 didn't)	11 (1 didn't)	25
<i>What would you change in activities?</i>						
<i>'Nothing' / 'not much'</i>	5	2	7	3	1	4
<i>Do speaking</i>		3	3	1	2	3
<i>More variety (skills, activity type, more fun/interest)</i>	5	3	8	4	3	7
<i>Shorter, fewer</i>	1		1	5		5
<i>Organisation (groups, teachers)</i>		3	3		3	3
<i>other</i>	2	3 (2 about EGI OHT, 1 too easy)	5	1 (too easy)	3 (2 "no opinion type [affective] activities"	4

Table 4.8 Summary of results from pupils' questionnaires regarding the PI and EI intervention materials

¹³ As some learners gave both positive and negative comments, these figures do not necessarily add up to the n for that group.

Question 1: Have you enjoyed the activities?

Slightly fewer EI learners enjoyed their activities than PI learners. However, this is mainly due to a difference in class A, where two of the reasons given by the EI learners were organisational rather than directly related to the EI activities and 3 learners in the PI group also commented that they found the activities boring. Note also that there was one more EI learner than PI in class A.

Question 2: Did you find the activities helpful and useful?

There were three fewer EI learners than PI learners in class B who reported they found the activities useful and helpful. Although this is a small number, it could indicate that EI did not give an equal sense of progression to both EI and PI learners in this school. However, two of these three learners were the pupils who had produced identical comments and both explained that they had not enjoyed being split up from their friends, which their behaviour had also demonstrated during the sessions. The reason they gave for not having found the activities helpful were that they found them too straightforward, yet both these learners achieved quite low scores in the tests. The third pupil was distressed during the questionnaire completion because teacher B, contrary to the agreed protocol, approached this pupil and read her responses. During the focus groups this pupil was very positive about the activities, giving detailed explanations about why she thought they were useful and had helped her. See appendix 37 for examples of responses to this question.

Question 3 Do you think the activities were any different from things you have done before? Underline: The activities were a bit different / weren't actually that different

There were a similar number of responses to these from both instructional groups. This was reassuring in that almost all learners felt that they had experienced something new, confirming the likelihood that any differences in performance were not due to perceived novelty. This finding also suggested that although the EI activities had been designed to imitate the input processing induced by 'typical' listening and reading activities, the learners actually reported that they thought these activities were different from anything they had done before. However, the reasons

given were very wide ranging - including many organisational reasons (such as working in smaller groups, increased time efficiency, the use of worksheets compared to the text book or working from the board). Nevertheless, a handful of responses, mainly from PI learners, were related to where their attention had been focussed in the activities (e.g. “on the little bits”).

Question 4: Would you like the teacher to continue using this style of activity? Yes / no

Why / why not?

There were four more EI learners than PI learners who did not wish to do any more such activities. This difference can be accounted for by the two learners who filled in the questionnaire together and wrote identical comments and two learners who implied that they would like to do more but just less intensely.

Question 5: What type of activities have you been doing and what is your opinion about them?

¹⁴Overall there were slightly more positive comments from the EI learners than PI learners. This is due to differences in class B, and may have been because the PI learners wrote fewer comments overall.

Question 6: Have the activities changed the way you notice French grammar at all? Yes a bit / no not really. If so please explain.

Roughly similar proportions of PI and EI learners in school A reported that the activities had helped them to notice the grammar more. In class B, four more PI learners than EI learners reported this. It is possible that this may indicate that the EI learners did not notice the target features (presented in the EGI) in the subsequent EI activities. Three learners from the EI group in class B expressed this in the focus group discussions, commenting that they would have liked some practice of the language items presented in the EGI in the activities that followed (implying that the flood of items in the EI was not salient and/or helpful to them). In addition, the EI learners in class B frequently commented on the amount of new words they had

¹⁴ Responses to the first part of question 5 were not analysed as they were not relevant to this thesis: they offer some insight into learners’ articulation of their awareness of the activities they undertake.

learnt (suggesting they were aware of the lexical / semantic focus of the EI activities). However, as the numbers are small, and the sample for the focus groups was self-selecting, this phenomenon (EI learners in class B reporting that they did not attend to the target verb inflections) cannot be claimed with any certainty at this stage. Nevertheless, these findings do suggest further investigation of the learners' awareness of what they pay attention to in the input would be worthwhile. See appendix 37 for examples of responses to this question.

Question 7: What would you change about these activities?

Table 4.8 shows that there was one type of response to this question that was different in the PI and EI groups in class A: five PI learners expressed the feeling that the activities became monotonous, compared to just one EI learner. However, this did not seem to have a negative impact on the learner outcomes (analysed in chapters 5 & 6).

In summary, no conclusive marked differences were found in the responses to the questionnaires from pupils who had experienced PI and those who had experienced EI. Both groups reported that they enjoyed the activities and that they thought the activities were 'new' and 'different'. Any differences found in more specific responses were either considered inconclusive, due to the nature of the comments, or, as will be seen in chapters 5 and 6, irrelevant, in that they did not correlate with the test results. Overall, it is suggested that any differences in the test scores at post and dp tests are unlikely to be due to motivational / attitudinal issues as a consequence of the PI and EI materials.

4.3.3 Interviews with the teachers involved with the experimental intervention (teachers A and B)

Semi-structured interviews (appendix 30) were carried out at the end of the intervention phase of the study. They are not analysed in detail, as one of their main functions was to discuss the administration of the experiment. However, three key issues discussed were essential to the interpretation of the test results and so are discussed briefly here. Before each finding from the interviews is presented, an explanation of its relevance to the thesis is given.

It is acknowledged that in these interviews the validity of the teachers' responses may have been compromised by the fact that I had designed the materials. However, both teachers offered positive and negative criticism about the materials, including suggestions about improving them, and this implies that they felt reasonably comfortable talking to me honestly about their attitude to the intervention.

- i) As explained in chapter 1, one of the reasons for comparing PI with EI was that the EI materials were designed to place similar input processing demands on learners to those placed by 'typical' reading and listening materials i.e. they required learners to attend to lexical items in the input or overall sentential meaning in order to carry out the task and any processing of verb inflections would have been incidental. Both teachers agreed that the EI activities encouraged similar language learning processes to typical reading and listening activities found in text books and classroom resources. They confirmed that they felt that the essential focus of the activities was similar to normal comprehension exercises in that it mainly focussed learners' attention on lexical items or 'key vocabulary', although other details of the activities were different e.g. the presentation of the materials, the length of the text and the lack of closed questions in some of the EI affective activities.
- ii) It was important to control and assess any preference teachers may have had for either EI or PI. As laid out in chapter 3, this was mainly done in three ways: by asking the teachers to follow the guidelines and materials provided with each unit of work, by recording the lessons and ensuring that there was no apparent bias in the teachers' attitude and by sharing the teaching of both EI and PI between myself and the teacher. In addition, the teachers were asked about their attitude in the interviews and were generally equally positive about both sets of materials. However, teacher B did express some reservations about the EI activities in that she would not have followed the EGI with the EI activities, but would have done production practice or more explanations of the grammar points. Nevertheless, she thought that the two types of approaches were both useful and had helped the learners, and despite some initial concerns, she had been happy doing both EI and PI. The recordings of the EI lessons that she taught do not suggest that she expressed any reservations about the

materials to the pupils and they show that she administered the activities according to the guidelines. Both teachers suggested that the EI activities might be more suitable for learning vocabulary and PI for practising grammar. Both teachers said that they would re-use some of the EI and PI activities.

iii) As one of the broader objectives of this study was to explore an innovative approach to grammar pedagogy (PI), it was of interest to assess whether the teachers felt that PI really was 'new' to them. In the interviews, both teachers reported that they thought that they had never seen or used activities like PI before this study, agreeing that PI placed input processing demands on learners which other listening and reading activities do not.

4.4 Summary of chapter 4

Section 4.1 provided descriptions of the lessons in classes A, B and C prior to, during and after the intervention. These suggested that there were some key differences between class B and the other two classes in terms of general classroom management, sequencing of activities and approach to grammar pedagogy. It was suggested that learners in class A, and probably also in class C, would be more likely to expect learning activities to be sequenced according to language (either grammatical or lexical themes). They would also probably expect to be asked about/tested on a small set of target language features often embedded in learning activities, verb inflections in particular, on a regular basis. Learners in class B would be more likely to expect class activities to be organised in broadly defined communicative tasks. They would probably expect a wide range of grammatical features to be raised at any point in lessons, and would not normally then expect these to be reviewed / practised proactively.

In terms of adherence to the experimental protocol between the pre and post test, there was negligible focus on the target forms outside the experimental intervention in any of the classes, therefore it is considered that any differences between EI and PI groups at post test are likely to be due to differential impacts of these instructional approaches. There was no evidence of FonF/S on the target forms between the post

and dp test in class B, therefore the scores at dp test can be ascribed to the experimental intervention. However, the other two classes experienced extensive FonF/S in the target features after the post tests and any gains at dp test may be due to these.

Section 4.2 showed that the pre test scores were statistically significantly lower in class B compared to classes A and C in school 1 in the listening, reading and writing measures. However, in the speaking measure, classes A and B had statistically similar pre test scores. Classes A and C had statistically similar scores at pre test in all measures. This suggests that class C's post and dp test scores will be a useful comparison for class A's scores. It was also suggested that as class B was probably from a different population at the outset, in terms of the interpretation and written production of verb inflections, this study will provide some understanding of differential impacts of PI and EI at different developmental stages in these measures. In addition, several supplementary analyses will be necessary to ensure that these initial differences did not threaten the validity of the study.

Section 4.3 suggested that during the study there were no important changes in different groups' and classes' ranking of French lessons in terms of enjoyment or difficulty. A few slight differences between PI and EI learners' opinions of the intervention materials were either considered inconclusive or do not correlate in an intuitive way with the results presented in chapters 5 and 6. There was, however, one difference that may be relevant to this thesis: some data suggested that EI learners in class B did not find it helpful or may not have noticed (in a general sense) that the materials contained a flood of the target items presented in the EGI in order to promote the learning of them. The interviews with teachers A and B suggested three main points: both teachers felt that EI materials did simulate the input processing demands of normal listening and reading activities carried out in MFLs classrooms; both teachers felt that the PI and EI materials were useful for the learners but that EI was perhaps more useful for teaching vocabulary and PI for teaching grammar; PI was an innovative approach to grammar pedagogy for the two teachers interviewed.

Chapter 5 Results and Analysis of the Interpretation Language Tests

Introduction

This chapter contains the results, analysis and discussion of the data from the aural and written interpretation tests (presented in sections 5.1 and 5.2 respectively). Both listening and reading measurements were used to investigate the possibility that the impact of PI depends on modality. As the research questions suggest (see section 2.5), analyses were carried out to assess the impact of the following independent variables on the language measures: time of test (pre, post and delayed post test), type of instruction (PI, EI, and Class C) and class (A, B or C). Section 3.8 described the analysis procedures used, thus there is limited explanation of procedures underlying statistics provided in chapters 5 and 6. It is emphasised that the procedures are used extensively in the social sciences and FonF/S and PI studies and detailed mathematical discussions of the procedures are not within the scope of this thesis. The pattern of analyses is similar across all measures and is described in section 5.1 and then repeated without further explanation throughout. Analysis is mainly quantitative, including descriptive, and both parametric and non-parametric inferential statistics. Parametric tests are used throughout in line with the tradition in FonF/S studies and PI studies. However, as discussed in chapter 3, due to violations of the assumptions required for certain tests, non-parametric tests were also deemed necessary in most cases.

Some interpretation of the results is given alongside their presentation, though each section ends with a summary which establishes preliminary implications for the research questions.

5.1 Analysis of the listening tests

The total of all listening test activities (including all target features and all pupils' data) was calculated as a percentage score from the total raw score of 48. A

percentage was used for all measures so that descriptive statistics are more easily compared (this was particularly necessary for the narrative production measures where the total scores depended on the individual learner).

5.1.1 *Can parametric tests be used?*

The statistics for all tests of the normality of the distributions are given in appendix 29 and a summary is given in the body of the thesis. The normality of distribution tests for the listening scores showed that most of the scores of the merged instructional groupings (PI and EI), the intact classes and EI and PI within classes A and B at pre, post and dp test tests are not statistically significantly different to a normal distribution of a sample with the same mean and standard deviation at the 95% confidence level. The only exceptions were the merged EI post test scores and the overall pre and post test results, which are not normally distributed at the 95% confidence level, but could all be considered normal at the 99% confidence level. The possible non-normality in the merged EI post test scores is due to a positive skew (i.e. more learners achieving low scores when compared to a normal distribution). As the majority of the relevant samples of data can be considered normally distributed (and given that the normality assumption can be slightly violated without compromising statistical accuracy), the listening scores could undergo parametric tests.¹

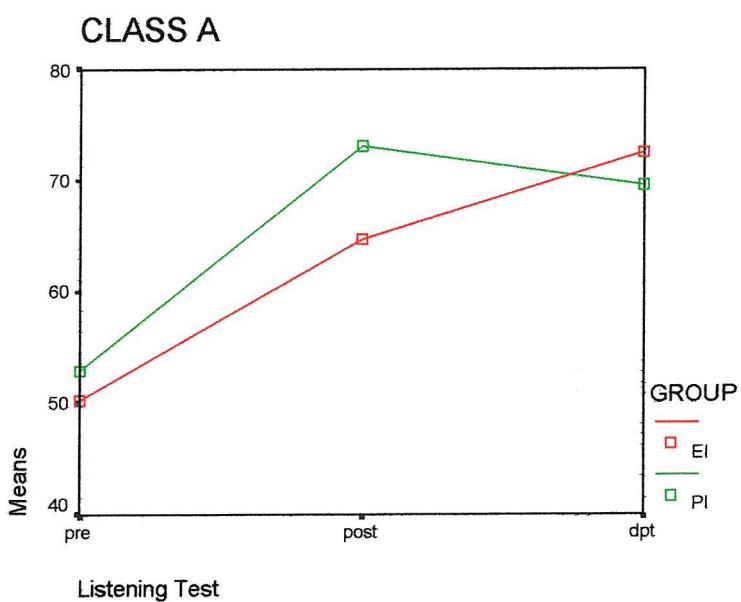
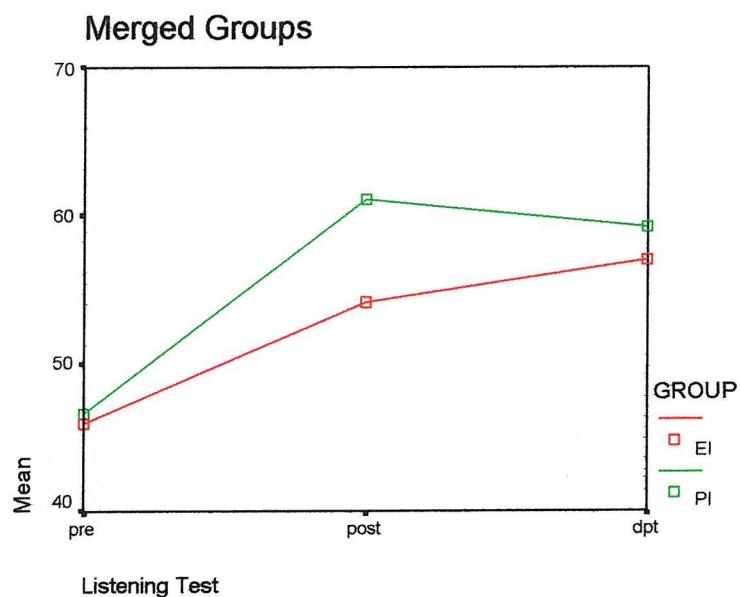
5.1.2 *Descriptive statistics*

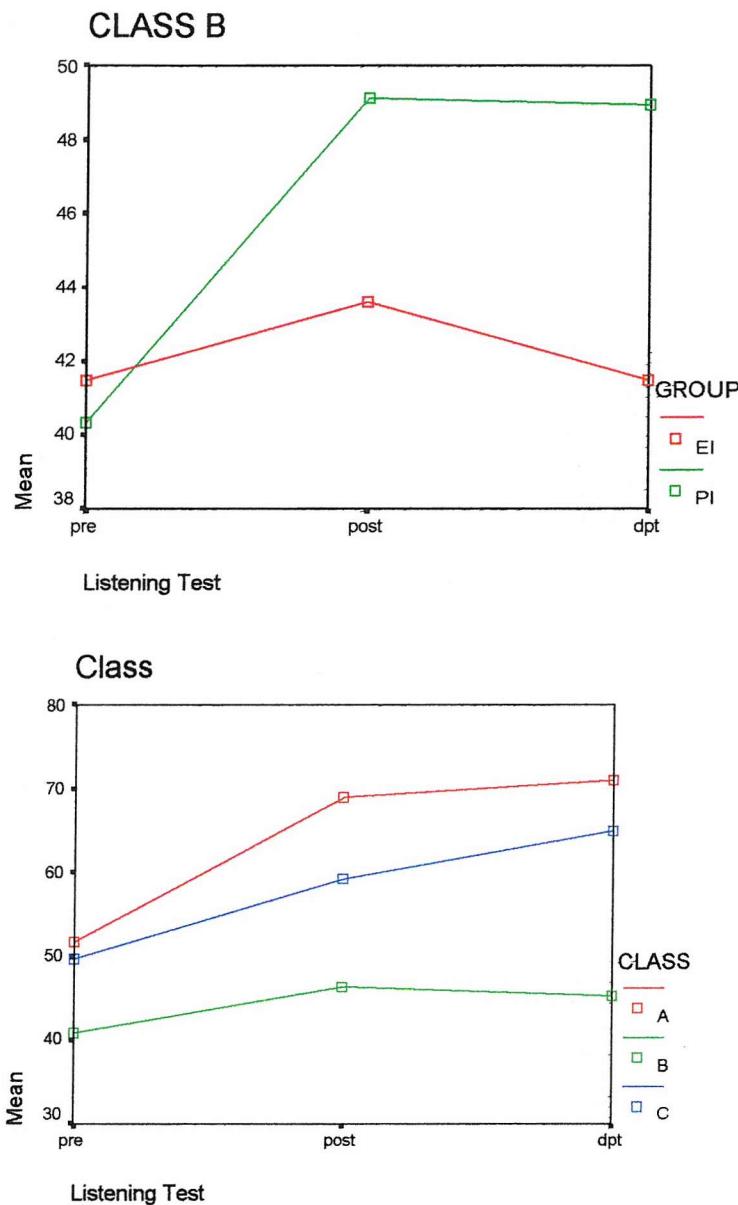
a) Graphs

Graphs showing the mean listening test scores² in the merged EI and PI groups, classes A, B and C, and PI and EI groups in classes A and B provide an overview of the scores e.g. to see the direction of the differences between the tests. This section also presents an analysis of the effect sizes of the gains made by particular groups. Descriptive statistics (the mean, standard deviation and n of each sample) for the total listening scores are given in appendix 31.

¹ Although non-parametric tests were also carried out to provide extra support for the results of the parametric tests, to ensure that the few slight deviations from normality did not affect the interpretation of the findings.

² The lines are not intended to imply any linear relationship in progression between the tests.





b) Effect sizes of gains

The effect sizes provided here give an indication of the magnitude of the mean gains made by PI compared to EI learners. As explained in section 3.8.7, mean *gains* between tests were used in order to eliminate slight differences between the groups at pre test (although these differences were not statistically significant, as shown in section 3.3 and appendix 15).

Section 3.8.7 also explained that 0.75 standard deviation units (sdu) is used as a benchmark, as found in Norris & Ortega (2000) when the comparison group was

exposed to and interacted with materials in which the target L2 form was embedded (as in the EI in this study).

Three effect sizes were calculated for each of the gains: a comparison of the effect of the merged PI group over the merged EI group and comparisons of the effect of the PI group over the EI group within classes A and B.

<i>Effect size of...</i>	<i>Pre- post gains</i>	<i>Post- dp gains</i>	<i>Pre - dp gains</i>
merged PI over merged EI	0.47	-0.30	0.07
class A PI over EI	0.49	-0.78	-0.48
class B PI over EI	0.59	0.13	0.53

Table 5.1 Effect sizes of gains between listening tests

There was a positive effect size of the PI merged group in the pre-post gains, although by the time of the dp test there was no effect size between the merged PI and EI groups. The reader is reminded that the results of the merged group scores at dp test probably reflect the fact that between the post and dp tests half the merged PI and EI groups (i.e. learners in class A) had FonF/S in the perfect tense (see section 4.1).

In class A, PI had an advantage in the pre-post gains. These gains were countered by those made by EI after the post test and, at dp test, the negative effect size suggests that the EI instruction may have had favourable effects over PI. However, as mentioned above, attributing gains between post and dp tests in class A to the experimental interventions cannot be conclusive due to the considerable FonF/S in this class after the post test.

In class B the advantage of PI over EI at post test was maintained at dp test, where the PI gains had an effect of about two thirds of the magnitude found by Norris & Ortega (2000) (0.75 sdu).

5.1.3 Repeated measures ANOVA to assess whether the instructional type and/or class had any statistically significant impact on the listening scores

To assess the impacts of instructional type (EI and PI) and class (A, B and C) on overall tendencies in the aural comprehension of verb inflections, a repeated measures analysis of variance (ANOVA) was conducted on the test scores, using SPSS version 11, with class (CLASS) and instructional group (GROUP) as between subject factors and the three listening test scores as within subject measures (LIST).

Levene's test of equality of error variances showed that the variances of some or all of the sample groupings within the post test listening scores cannot be considered equal (appendix 32). However, transforming the scores using the square root did render the variances equal (see section 3.8.5). The subsequent analyses of the transformed scores (with equal variances) produced the same results, in terms of statistical significance, as the analysis of the non-transformed scores (with unequal variances). The analysis of the non-transformed scores is presented here for ease of interpretation.

The ANOVA (table 5.2) showed that there was some variance between samples that was statistically significantly different from the variance within each sample, showing a significant effect for the time of the test (pre, post, dp) and for the class (A, B, C), though not for the instructional type (GROUP: merged PI, merged EI or class C). The interaction between the class and group was only approaching statistical significance at the 95% confidence level, though it was significant at the 90% level, and may indicate a tendency for the instructional type to have different impacts depending on the class learners were in.

Source	Type III Sum of Squares	df	Mean Square	F	Sig. ³
LIST	7967.829	2	3983.914	47.088	.000
LIST * GROUP	285.360	2	142.680	1.686	.188
LIST * CLASS	1721.868	2	860.934	10.176	.000
LIST * GROUP * CLASS	435.766	2	217.883	2.575	.079

Table 5.2 *Repeated measures ANOVA of the listening test results*

ANOVAs do not suggest where differences are in terms of the repeated dependent variable (between pre and post, pre and dp and/or post and dp tests), nor in terms of the different classes (CLASS) or instructional types (GROUP). For this, planned contrasts were necessary (see section 3.8.5 for a description). The results are shown in table 5.3.

Source	LISTTEST	Type III Sum of Squares	df	Mean Square	F	Sig.
LIST	Post vs. Pre	9810.359	1	9810.359	76.438	.000
	Dp vs. Pre	13759.873	1	13759.873	67.275	.000
	Post vs. Dp	333.253	1	333.253	1.907	.171
LIST * GROUP	Post vs. Pre	523.794	1	523.794	4.081	.047
	Dp vs. Pre	30.368	1	30.368	.148	.701
	Post vs. Dp	301.919	1	301.919	1.728	.192
LIST * CLASS	Post vs. Pre	1922.388	1	1922.388	14.978	.000
	Dp vs. Pre	3102.642	1	3102.642	15.169	.000
	Post vs. Dp	140.572	1	140.572	.804	.372
LIST * GROUP * CLASS	Post vs. Pre	3.728	1	3.728	.029	.865
	Dp vs. Pre	701.040	1	701.040	3.428	.068
	Post vs. Dp	602.530	1	602.530	3.448	.067

Table 5.3 *Planned contrasts of listening tests*

These planned contrasts found that the differences between the tests found by the ANOVA were due to differences between the pre and post, and between the pre and dp test listening scores.

They also found statistically significant differences due to class effect between pre and post and between pre and dp test.

³ Mauchly's test showed that this data does not violate the assumption of sphericity (see section 3.8.5) (Mauchly's W statistic 0.931, Chi-square 5.619, df 2, p=0.060 p>0.05) and so the F statistic can be

Although the ANOVA did not find statistically significant effects for the interaction LIST*GROUP - the effect of instructional type on the scores - the planned contrasts imply there may have been a tendency for type of instruction to have an effect between pre and post test, but not between the other tests.

The contrasts may also imply that a combination of class AND type of instruction may have impacted on learners' scores between post and dp test and between pre and dp test. However, these planned contrasts cannot be considered completely reliable as the ANOVA did not find overall significance in those effects. Nevertheless, this tendency was worth further investigation.

As the research aim was to discover the impacts of PI compared to EI, the analyses were also carried out without class C, as this 'instructional group'⁴ could have affected the findings for the interactions test*group and test*class*group. The results of the analysis were, in general, no different to those from the analysis including all the data, and are therefore not shown here though they are included in appendix 33⁵.

As there were differences between some of the samples at pre test (some of which were statistically significant and others which were not) an analysis using the gains between the tests also supported the findings above (appendix 34).

5.1.4 Separate analyses of each class

The above tests provided some evidence that instructional type had an effect in one class but not in another. Chapter 4 had previously suggested that class B was probably from a different population. Therefore, separate repeated measures ANOVAs were carried out on each class i.e. as though mini-studies had been carried

considered valid.

⁴ For the analysis in SPSS, class C was coded as an intact 'class' and as an instructional 'group', so as to compare the merged PI and EI groups with class C as a non-active control.

⁵ The only slight difference was that the planned contrasts between the pre and post tests for the interaction list*group found slightly different results: $F=4.032, 1, p=0.050$ (rather than $p=0.47$ when class C was included). This is not considered sufficiently different to cast doubt on the existence of tendencies due to the different instructional types.

out on different populations. Planned contrasts (simple and repeated) were carried out in order to investigate the source of any differences found by the ANOVA.

For class A, Levene's test of equality of error variances showed that the variances within the pre and post tests were homogenous at the 95% confidence level, though at dp test only at the 99% confidence level. In class B, the variances in the pre and dp test variables were homogenous at the 95% confidence level, but at post test only at the 99% confidence level (appendix 32). The results of the ANOVAs must again be interpreted with some caution.

	Source	Type III Sum of Squares	df	Mean Square	F	Sig. ⁶
Class A	LISTTEST	6314.456	2	3157.228	43.386	.000
	LISTTEST * GROUP	447.819	2	223.910	3.077	.055
Class B	LISTTEST	443.785	2	221.892	2.100	.133
	LISTTEST * GROUP	276.539	2	138.269	1.309	.279
Class C	LISTTEST	3548.351	2	1774.176	23.019	.000

Table 5.4 Three separate repeated measures ANOVAs on separate classes for listening tests

	Source	LISTTEST	Type III Sum of Squares	df	Mean Square	F	Sig.
Class A	LISTTEST	Post vs. Pre	8343.484	1	8343.484	63.582	.000
		Dp vs. Pre	10478.413	1	10478.413	74.383	.000
		Post vs. Dp	121.472	1	121.472	.738	.398
	LISTTEST * GROUP	Post vs. Pre	223.797	1	223.797	1.705	.203
		Dp vs. Pre	224.023	1	224.023	1.590	.218
		Post vs. Dp	895.639	1	895.639	5.444	.028
	LISTTEST	Post vs. Pre	795.563	1	795.563	6.189	.020
		Dp vs. Pre	502.198	1	502.198	1.838	.187
		Post vs. Dp	33.593	1	33.593	.145	.707
	LISTTEST * GROUP	Post vs. Pre	302.245	1	302.245	2.351	.138
		Dp vs. Pre	502.138	1	502.138	1.838	.187
		Post vs. Dp	25.233	1	25.233	.109	.744
Class C	LISTTEST	Pre vs. Post	2755.592	1	2755.592	21.940	.000
		Post vs. dp	949.444	1	949.444	7.061	.013
		Dp vs. Pre	6940.019	1	6940.019	34.290	.000

Table 5.5 Planned contrasts in each separate class for listening tests

⁶ Mauchly's test of sphericity showed that this assumption was upheld (.982, .466, 2, p=.792)

In class A the repeated measures ANOVA suggested that there were statistically significant differences between the different tests. The contrasts suggested that the differences were between the pre and post and between the pre and dp test. The ANOVA indicated that the effect of instructional group was approaching statistical significance ($p=0.055$), but the planned contrasts showed that this was between the post and dp test only, despite the fact that there was no differential experimental intervention for the PI and EI groups between these tests. The graphs in section 5.1.2 show that this is due to the increase in the EI group's scores between the post and dp test and the slight decrease in the PI group's scores.

The intact class B, made statistically significant gains, at the 95% confidence level, between pre and post tests, as shown in table 5.5. The effect of instructional group was not statistically significant, though the planned contrasts indicated that it may be worth investigating further whether there was a tendency for instructional group to affect the scores between pre and post tests.

It was found for class C's scores that there were significant differences between all pairs of the tests i.e. that class C improved from pre to post and from post to dp test.

5.1.5 Non-parametric tests

There are several reservations about the analyses presented above due to the lack of homogeneity of variances and the borderline normality of some of the sample distributions. Further exploration of the data was carried out using non-parametric tests (described in section 3.8.6). The results were identical to those laid out above, in terms of findings of statistical significance, therefore the statistical details are not presented here but can be found in appendix 27. A summary is given here.

It was found that for the whole set of learners' scores, there were significant differences between the pre and post and between pre and dp test but not between post and dp test. This was also the case for the intact class A. In class C there were statistically significant differences between all tests, including the post and dp test. In the intact class B there were no statistically significant differences between any of

the tests, though the Wilcoxon paired comparison suggested there may have been tendencies for learners to improve their scores between pre and post tests.

In both the merged PI and EI groups there were significant differences between pre and post and between pre and dp tests only. Within class A, both the EI and the PI groups showed statistically significant improvement between the pre and post tests and between the pre and dp test. Only the EI group showed a difference between the post and dp test. Within class B, neither the EI nor the PI group made statistically significant gains between any of the tests according to the Friedman test. However, the Wilcoxon paired comparisons suggested that there were tendencies for the PI learners to improve their scores between pre and post ($p<0.05$) and maintain this gain at dp test ($p=0.058$). However, as the Friedman did not find statistically significant differences between all the tests, it is acknowledged that the results of these Wilcoxon tests does not take account of a slightly increased chance of making a type 1 error (asserting there is a difference where there is none, Field 2000 pp243-4).

5.1.6 Assessing the possibility that learners randomly selected responses in the listening tests.

All the listening test activities required the learners to select responses. As with any multiple choice measure, there is a possibility that chance played a part in the scores. This type of analysis has not been carried out in previous PI studies, despite the fact that random selection was possible in the interpretation tests used. It is useful to determine whether there were any patterns in the extent of random selection in terms of specific class or instructional groups, particularly to explore how any such patterns related to the previous analyses. In addition, this section is intended to provide a broad indication as to whether the test format (multiple choice) should be taken into account when interpreting the results.

One sample t-tests were carried out to determine if there was any difference between the observed score and that which could be obtained by random selection. The null hypothesis was that there was no statistically significant difference between the learners' scores and scores obtainable by chance selection and the alternative hypothesis was that the learners' scores were statistically different from those

possible using chance selection. One sample t tests were carried out at each of pre, post and dp test. The 'test statistic' against which the observed data were compared was the score that learners could have achieved by selecting their response to each item entirely randomly⁷. Results are shown in table 5.6.

It is acknowledged that this analysis cannot comment on the target-likeness of the system that learners were using e.g. they may have used a system for interpreting aural verb endings that happened to result in target-like behaviour, or, where the statistics suggest that 'random selection' was occurring, this does not exclude the possibility that learners were indeed using some sort of 'system' in their interpretation. The words 'random' and 'chance' are used to refer to statistical probability, and not to suggest a lack of any learner system.

If learners were randomly assigning responses throughout the four listening tasks, they could be expected to achieve 42.71% - this was therefore the test statistic with which the observed values were compared. (Analyses were carried out on each of the four tasks individually. Because similar patterns emerged as the analysis of the total listening scores, only the latter are presented here).

⁷ As few learners used the category of 'not sure' in the multiple choices, this was not counted as a possible response i.e. in an item where learners had 4 choices plus 'not sure', it was considered that they could score 25% by using random selection (rather than 20% if the 'not sure' category was included - assuming that ticking 'not sure' meant that when they did choose an actual language response, it was not random). This means that this analysis may slightly over-estimate the extent to which random selection was used.

<i>Group/class</i>		<i>Mean</i>	<i>St dev.</i>	<i>T statistic</i>	<i>Df (n-1)</i>	<i>P value</i>
<i>Class A</i>	<i>pre</i>	51.64	16.80	2.811	27	.009
	<i>post</i>	68.53	16.06	8.657	28	.000
	<i>dp</i>	71.05	17.16	8.895	28	.000
<i>Class B</i>	<i>pre</i>	40.89	9.93	-.950	26	.351
	<i>post</i>	46.45	9.21	2.112	26	.044
	<i>dp</i>	46.45	9.21	2.112	26	.044
<i>Class C</i>	<i>pre</i>	49.72	12.62	3.044	29	.005
	<i>post</i>	59.31	13.87	6.553	29	.000
	<i>dp</i>	64.93	13.80	8.821	29	.000
<i>EI group</i>	<i>pre</i>	46.06	14.73	1.183	26	.248
	<i>post</i>	54.55	16.04	3.835	26	.001
	<i>dp</i>	58.57	18.20	4.526	26	.000
<i>PI group</i>	<i>pre</i>	46.65	15.06	1.385	27	.177
	<i>post</i>	60.99	17.90	5.499	28	.000
	<i>dp</i>	59.77	19.15	4.797	28	.000
<i>Class A EI</i>	<i>pre</i>	50.30	16.78	1.692	13	.115
	<i>post</i>	64.73	15.88	5.190	13	.000
	<i>dp</i>	72.47	14.11	7.891	13	.000
<i>Class A PI</i>	<i>pre</i>	52.98	17.35	2.214	13	.045
	<i>post</i>	72.08	15.94	7.136	14	.000
	<i>dp</i>	69.72	20.00	5.232	14	.000
<i>Class B EI</i>	<i>pre</i>	41.51	11.04	-.393	12	.701
	<i>post</i>	43.59	5.80	.547	12	.594
	<i>dp</i>	43.59	5.80	.547	12	.594
<i>Class B PI</i>	<i>pre</i>	40.33	9.16	-.973	13	.348
	<i>post</i>	49.11	11.08	2.161	13	.050
	<i>dp</i>	49.11	11.08	2.161	13	.050

Table 5.6 One sample t tests to explore the possibility that learners were using random selection in the listening test.

As would be expected, this analysis reflects the findings of the analysis of the impact of class and instruction on the total listening scores. Classes A and C began with some kind of system that indicates a 'better than random' interpretation of verb inflections, and this was maintained at post and dp test. Class B began with scores that were not significantly different to random selection, but at post and dp test there were statistically significant differences between their scores and those possible by random selection. Both the merged PI and EI groups and the EI group in class A had pretest scores that were suggestive of random selection, though post and dp test scores suggest that learners were selecting responses systematically. The PI group in class A at pretest obtained scores that are unlikely to be due to random selection, and

this is even clearer at post and dp tests. In class B, the PI group began with scores which were no different to random selection but at post and dp test their scores were significantly different ($p=0.050$) to those obtainable by random selection. The EI group in class B achieved scores obtainable by random selection throughout the study.

5.1.7 Discussion and further analysis of class C – the non-active control

The analyses above suggested that PI and EI may have had differential effects depending on the class learners were in: in class A, both groups seem to have made equivalent gains (and maintained these after approximately 14 weeks); in class B, only the PI learners made gains (and maintained them after approximately 14 weeks, despite having no FonF/S between post and dp tests). Some explanation of these apparently contradictory findings was required.

These findings could constitute evidence in support of IP principles 1a-e. PI learners in class B (and also possibly in class A) improved their ability to interpret and produce verb inflections after they had carried out PI activities which highlighted the meanings for person, number and tense that the verb inflections can carry. In class B, EI learners seemed unable to detect verb inflections whilst also interpreting items of higher CV (i.e. lexical items or overall sentential meaning), as would have been required in the EI activities for learning gains to have been made. In contrast, EI learners in class A may have been able to detect the target verb inflections *incidentally* to the items of higher CV because they were at a higher developmental stage in terms of interpretation and production of verb inflections (and possibly in terms of metalinguistic knowledge and vocabulary range⁸) than learners in class B (see section 4.2). The reader is reminded that IP sub-principle 1e) is “The Availability of Resources Principle: For learners to process either redundant meaningful grammatical forms or nonmeaningful forms, the processing of overall sentential meaning must not drain available processing resources”. That is, learners in class A may have had sufficient resources to detect items of low CV whilst also interpreting lexical items and/or overall sentential meaning.

A further possible explanation is also a consequence of classes A and B coming from different schools. Two analyses above (the repeated measures ANOVA carried out on separate classes and the equivalent non-parametric tests, see tables 5.4 and 5.5 and appendix 27) suggested that class C, the non-active control, made statistically significant gains between pre and post tests. Yet, the reader is reminded, between pre and post test class C learners had very limited, if any, exposure to FonF/S instruction (section 4.1). Therefore, their improvement between pre and post tests may be evidence of a test effect. There is also a possibility that this test effect may have affected learners in class A, given that, as mentioned previously, class C was parallel to class A in the same school, had had similar prior instruction in the target features and achieved statistically similar scores at pretest in all measures (see section 4.2). It is possible that given the ethos of teaching and testing grammar at this school (see section 4.1), the pre and post tests may have reactivated some language for learners in school 1 (i.e. including the EI learners in class A) enabling them to make gains between pre and post test.

However, although class C made statistically significant gains between the pre and post test, the graphs in section 5.1.2 suggested that these gains were not as large as the gains made by class A. By comparing the progression made by classes C and A between pre and post test, this possible explanation for the class A's EI learners' gains between pre and post tests was explored. It was found that the effect size of class A's gains over class C's gains between pre and post test was 0.67 sdu (though this is slightly lower than the mean found by Norris & Ortega (2000), 0.96 sdu when the comparison group had no exposure to the target feature). Independent samples t tests were also used to compare the pre-post test gains made by classes A and C (see table 5.7). (As there was no statistically significant difference between the gains made by the EI and PI group in class A, the whole of class A was compared to class C, as this would indicate the extent of the impact of the school / test effect.)

⁸ based on data from an action research project and lesson observations for the current study

t	df ⁹	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
					Lower	Upper
2.563	56	.013	7.67814	2.99561	1.67722	13.67906

Table 5.7 *Independent samples t test to compare the gains made by classes A and C between the pre and post listening tests*

It was found that between pre and post test Class A's gains were statistically significantly higher than class C's i.e. although class C made some gains between the pre and post test without any FonF/S instruction on the target form, they did not make as much progress as class A. This implies that the gains made by EI learners in class A cannot be entirely explained by the test effect (which, in turn, was probably due to the ethos of teach and test grammar in school 1). It is likely that EI learners in this class also benefited in some way from the experimental instruction, possibly by processing verb inflections incidentally to the EI tasks (as discussed above). However, it should be acknowledged that the EI (and PI learners) in class A may have benefited from the EGI at the start of each teaching unit, particularly perhaps given that these learners were accustomed to explicit grammar instruction (section 4.1)¹⁰. However, as discussed in section 2.3.1 and shown in, for example, VanPatten & Oikkenon 1996 and Benati 2004, it is unlikely that this brief explicit instruction alone had any significant impact on learning. It remains a possibility that the combination of the EGI with the EI may have benefited the EI learners in class A.

5.1.8 *Summary of 5.1: Analysis of the listening tests*

This section has presented a range of analyses of the listening scores. These analyses have all indicated that in the data set as a whole there were statistically significant differences between the pre and post test and between the pre and dp test. All analyses found that learners' scores depended on the class they were in (regardless of instructional type) between pre and post and between pre and dp tests.

⁹ Equal variances of the data could be assumed.

¹⁰ A similar explanation for gains made by class B learners can be eliminated as the EI learners made no gains in this class.

The instructional type alone (i.e. regardless of class learners were in) did not make a statistically significant impact on the scores.

Most of the analyses suggested that there may be a tendency for these effects to interact i.e. the class and instructional type together influenced the scores. In class A there is no evidence to suggest that PI had any advantages over EI as both groups made significant gains and maintained them at dp test. According to all analyses, the PI group in class B made gains approaching statistical significance between the pre and post tests and, according to the non-parametric test, their gains between pre and dp test were also approaching statistical significance (though this cannot be considered conclusive). The EI group in class B did not make any significant gains between any tests and neither group made gains after the post test.

Explanations for the different impact of EI in classes A and B were offered. It was suggested that the EI learners in class A may have benefited from the input flood (and possibly the EGI) in the EI because they were at a higher developmental stage than learners in class B and so had sufficient processing resources available to detect verb inflections incidentally to lexical items and/or sentential meaning. It was noted that a school effect (the background ethos of teach and test grammar) can not entirely explain the gains made by EI learners in class A as they made greater gains than equivalent learners in the non-active control (class C).

The analysis of whether learners had selected responses randomly suggested that EI learners in class B were the only group who may have used random selection in all tests. PI learners in class B seemed to begin with a random selection process though they had developed a more systematic interpretation of verb inflections at post and dp tests. In general, learners in school 1 (classes A and C) began with a selection process that indicated they were already interpreting verb inflections systematically, and this was maintained at post and dp tests.

5.2 Analysis of the reading tests

The analysis in this section follows a similar structure to the analysis of the listening scores¹¹.

The scores from all the written interpretation activities were added together and calculated as a percentage out of a raw score of 55.

5.2.1 *Can parametric tests be used?*

To assess whether the data was normally distributed, Kolmogorov-Smirnov (K-S) and Shapiro-Wilks (S-W) statistics were calculated for each of the samples within the pre, post and delayed tests (see appendix 29).

In classes A and C all of the reading tests produced data which was not statistically significantly different to a normal distribution of a sample with the same mean and standard deviation as the data obtained. Class B in the pre test has scores that, according to the K-S test could be statistically significantly different from a normal distribution at the 95% confidence level, though not at 99% and not according to the S-W statistic. Class B post and dp test scores can be considered normally distributed. In terms of the merged instructional groups, one data set (post PI) can be considered normally distributed. Four others could be considered normally distributed at the 99% level, but not at the 95% confidence level. One of these (PI, dp test) has a distribution approaching normality at the 95% level and for another (pre test EI) the W-S statistic indicated a normal distribution. The merged EI dp test should be considered non-normally distributed.

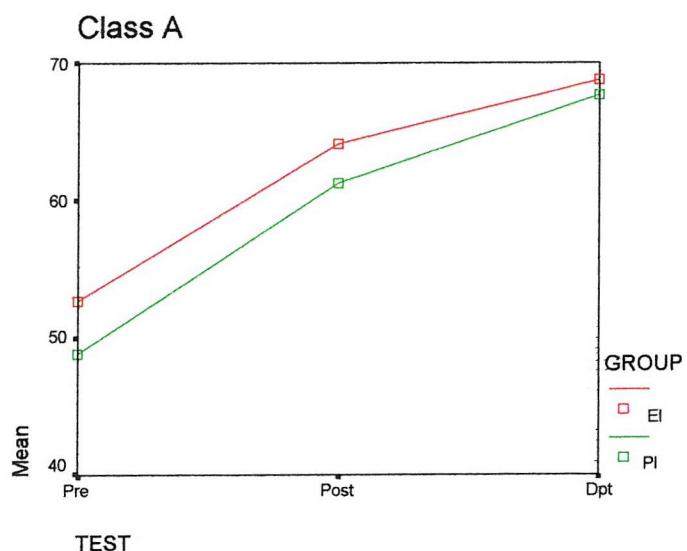
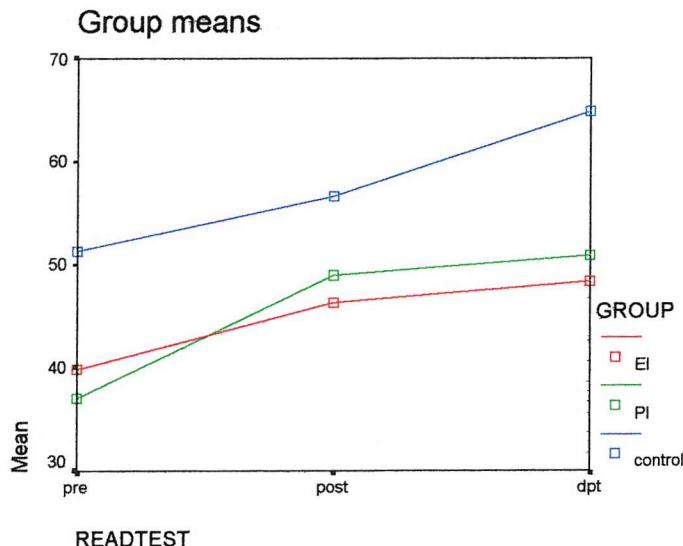
As the majority of the reading test data is normally distributed, and as at a 99% confidence level *all* the data can be considered normally distributed, parametric tests were carried out on the scores, in line with previous PI studies. However, caution is

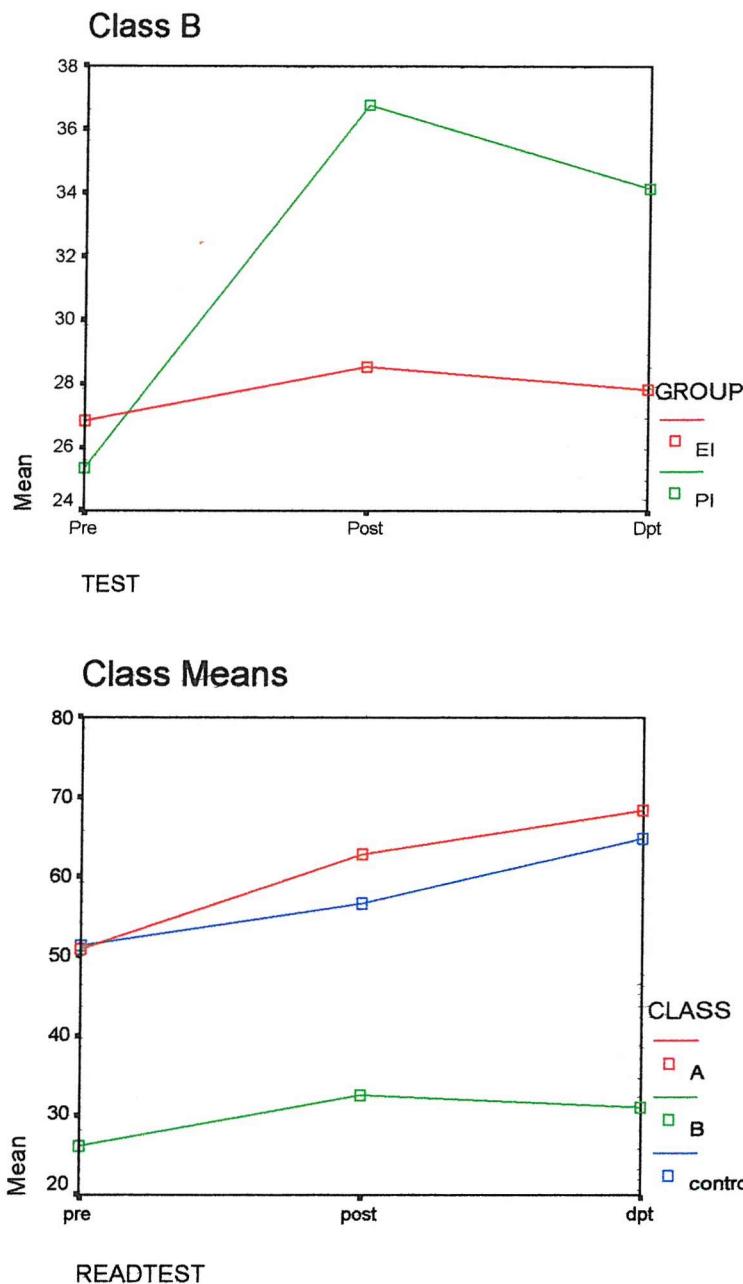
¹¹ An analysis of the role of random selection in one of the reading tasks (interpretation of tense, with two possible responses) produced very similar results to those presented in the analysis of the listening tasks in section 5.1.6. As most of the reading activities had more than 6 possible responses, further analysis of this kind was not considered necessary.

required in the interpretation of these tests and non-parametric tests were also carried out.

5.2.2 Descriptive statistics

a) Graphs





b) Effect sizes on gains between tests

<i>Effect size of...</i>	<i>Pre- post gains</i>	<i>Post- dp gains</i>	<i>Pre - dp gains</i>
merged PI over merged EI	0.48	0.03	0.38
class A PI over EI	0.10	0.25	0.21
class B PI over EI	0.93	-0.31	0.78

Table 5.8 Effect sizes of gains between reading tests

At post test the merged PI group had an effect size of slightly over half the 0.75 sdu benchmark (for comparing FonF/S approaches to instruction where learners interacted with materials in which the target form was embedded, see section 3.8.7). Although the merged PI group appeared to maintain some effect size over the merged EI group at dp test, the reader is reminded that between the post and dp tests half the merged PI and EI groups (i.e. class A learners) had FonF/S on the target features in this study.

In class A, neither instructional group had a clear advantage in terms of the effect size of their gains at post test. (The slightly favourable impact of PI on post-dp test gains can not be attributed conclusively to the experimental intervention, as explained above).

In class B the effect size of PI gains was even of the magnitude found by Norris & Ortega (2000) when the comparison group had *no* exposure to the target features (0.96 sdu). By dp test, the PI effect size was very similar to the 0.75 sdu benchmark.

5.2.3 Repeated measures ANOVA to assess whether the instructional type and/or class had any statistically significant impact on the reading scores

As with the listening test scores, a repeated measures ANOVA and planned contrasts were carried out on the scores from all 3 classes, with class and instructional group as between subject factors and the three test scores as within subject measures.

Levene's test for equality of variance showed that this assumption does not hold for this data (appendix 32). This means that the results of the repeated measures ANOVA cannot be taken as firm statistical 'proof', but can be considered as an indication of trends. However, as in the analysis of the listening scores, the transformation of the data using the square root achieved equal variances. As both analyses coincided in their main implications, only the analysis of the non-transformed scores is presented for ease of interpretation, though some aspects of the analysis of the transformed scores are discussed.

Source	Type III Sum of Squares	df	Mean Square	F	Sig. ¹²
READ	6161.136	2	3080.568	60.548	.000
READ * GROUP	261.488	2	130.744	2.570	.080
READ * CLASS	1091.007	2	545.504	10.722	.000
READ * GROUP * CLASS	131.759	2	65.880	1.295	.277

Table 5.9 Repeated measures ANOVA for the reading scores

Source	READTEST	Type III Sum of Squares	df	Mean Square	F	Sig.
READ	Post vs. Pre	5511.713	1	5511.713	54.588	.000
	Dp vs. Pre	11791.825	1	11791.825	94.646	.000
	Post vs. Dp	1179.872	1	1179.872	14.802	.000
READ * GROUP	Post vs. Pre	399.918	1	399.918	3.961	.050
	Dp vs. Pre	384.393	1	384.393	3.085	.083
	Post vs. Dp	.154	1	.154	.002	.965
READ * CLASS	Post vs. Pre	399.812	1	399.812	3.960	.050
	Dp vs. Pre	2167.638	1	2167.638	17.398	.000
	Post vs. Dp	705.571	1	705.571	8.851	.004
READ * GROUP * CLASS	Post vs. Pre	260.784	1	260.784	2.583	.112
	Dp vs. Pre	90.375	1	90.375	.725	.397
	Post vs. Dp	44.119	1	44.119	.553	.459

Table 5.10 Planned contrasts of reading tests

These results suggest that there were statistically significant differences between the three test scores. The interaction between the test scores and class has a statistically significant impact between post and dp test, when the intact classes resumed their normal lessons. Its effect between pre and post test is also apparent, though perhaps less clear cut ($p=0.05$).

The impact of instruction alone on the test scores was not statistically significant ($p = 0.08$), although the planned contrasts showed that there may be some tendency for this interaction to be more influential between pre and post tests. Furthermore, once the scores were transformed and had homogenous variances, the analysis showed that the interaction $\sqrt{\text{read}} * \text{group}$ was statistically significant (table 5.11), and the

¹² Mauchly's test of sphericity showed that this assumption is upheld with the reading scores (0.935, Chi-square 5.300, 2, $p>0.05$ ($=0.071$)).

planned contrasts (table 5.12) suggested this was between pre and post tests and between pre and dp tests.

The interaction between test*group*class (table 5.9) was not statistically significant, suggesting that learners' reading scores did not depend on the type of instruction AND the class they were in. Nevertheless, the planned contrasts (table 5.10) provided some indication that between pre and post tests both the class and instructional group had a tendency to influence learners' scores. This was also suggested by the planned contrasts of the transformed scores (table 5.12), though this is tentative as the ANOVA did not find statistical significance.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
√READ * GROUP	1.966	1.791	1.098	3.473	.039

Table 5.11 Extract from repeated measures ANOVA showing the effect of group on the transformed reading scores

Source	√READTEST	Type III Sum of Squares	df	Mean Square	F	Sig.
√READ * GROUP	Post vs. Pre	2.949	1	2.949	4.899	.030
	Dp vs. Pre	2.951	1	2.951	4.138	.045
	Post vs. Dp	2.618E-07	1	2.618E-07	.000	.999
√READ * GROUP * CLASS	Post vs. Pre	2.114	1	2.114	3.512	.065

Table 5.12 Extract from the planned contrasts on the transformed reading scores

When the analyses above were carried out excluding class C (as this 'instructional group' may have influenced the findings, rather than the PI and EI experimental groups), the findings were similar to those presented (see appendix 33). Therefore, differences involving 'group' were due to differential impacts of PI and EI, not class C.

As there were differences between some of the samples at pre test (some of which were statistically significant and others which were not) an analysis using the gains between the tests also supported the findings above (see appendix 34).

5.2.4 Separate repeated measures ANOVA on each class

As the analysis above has suggested that the effect of the instructional type may have had a different effect depending on the class learners were in, and as section 4.2 suggested that class B was from a different population, separate repeated measures ANOVAs were carried out on each class. Levene's test of equality of error variances showed that the variances were equal for all the variables in each class (appendix 32).

Class	Source	Type III Sum of Squares	df	Mean Square	F	Sig. ¹³
Class A	READ	4464.818	2	2232.409	37.428	.000
	READ * GROUP	26.479	2	13.240	.222	.802
Class B	READ	626.718	1.524	411.124	7.447	.004
	READ * GROUP	360.467	1.524	236.465	4.283	.030
Class C	READ	2777.029	2	1388.515	27.439	.000

Table 5.13 Three repeated measures ANOVAs on the separate classes for the reading tests

Class	Source	READTEST	Type III Sum of Squares	df	Mean Square	F	Sig.
Class A	READ	Post vs. Pre	3997.514	1	3997.514	37.580	.000
		Dp vs. Pre	8543.878	1	8543.878	50.710	.000
		Post vs. Dp	853.061	1	853.061	10.276	.004
	READ * GROUP	Post vs. Pre	7.550	1	7.550	.071	.792
		Dp vs. Pre	51.979	1	51.979	.309	.583
		Post vs. Dp	19.908	1	19.908	.240	.628
Class B	READ	Post vs. Pre	1158.002	1	1158.002	10.231	.004
		Dp vs. Pre	648.971	1	648.971	6.384	.018
		Post vs. Dp	73.180	1	73.180	1.945	.175
	READ * GROUP	Post vs. Pre	641.196	1	641.196	5.665	.025
		Dp vs. Pre	415.922	1	415.922	4.091	.054
		Post vs. Dp	24.282	1	24.282	.645	.429
Class C	READ	Post vs. Pre	872.749	1	872.749	10.197	.003
		Dp vs. Pre	5479.657	1	5479.657	52.188	.000
		Post vs. Dp	1978.682	1	1978.682	17.505	.000

Table 5.14 Planned contrasts in each separate class for the reading tests

¹³ As sphericity can be assumed (.817, 5.048, 2, $p>0.05$ ($=0.80$)) the values were the unadjusted values produced by SPSS.

For class A it was found that there were statistically significant differences between all the tests. However, the ANOVA suggested that in class A the type of input instruction did not have a statistically significant effect on the scores.

In class B the ANOVA suggested that there are significant effects for test: the planned simple and repeated contrasts suggested that there were differences between pre and post test and between the pre and dp test. The interaction readtest*group (i.e. that scores depended on the group the learners were in) was found to be statistically significant, the contrasts suggesting this was due to differences between pre and post. The differences were approaching significance between pre and dp test, but not between post and dp test. This suggests that after the intervention had finished there were no further significant gains but the beneficial effects of PI were maintained to some extent at dp test.

In class C the tests of within-subjects contrasts showed that there were statistically significant differences between pre and post, post and dp test and between pre and dp test.

5.2.5 Non-parametric tests to asses the impact of instruction and / or class on the reading scores

As the results of the non-parametric tests supported those of the parametric tests above, a summary of the findings is provided in this section and the statistics are given in appendix 27. The Friedman and Wilcoxon tests revealed that in the dataset as a whole there were significant differences between all pairs of the tests. This was also the case in classes A and C. The intact class B and the merged PI and EI groups made statistically significant gains between pre and post test and between pre and dp test, but there were no significant differences between the post and dp test scores.

Within class A both the EI and PI groups made significant gains between all pairs of tests. Within class B only the PI group made gains between the pre and post and between the pre and dp test. Neither group's scores changed significantly between the post and dp tests.

5.2.6 Discussion and analysis of class C, the non-active control

As discussed in section 5.1.7, the analysis of the reading tests has suggested that EI had a positive effect on learning in class A but no effect in class B. As class C (the non-active control in school 1) made gains between pre and post reading tests (as shown in the analyses above), despite having had no FonF/S, it is possible that the EI (and PI) learners in class A also made gains as a result of the same factors (probably the background ethos of teach and test grammar at school 1). However, the graphs in section 5.2.2 suggested that class A made greater gains than class C between pre and post test. In addition, the effect size of the gains made by class A compared to class C between pre and post tests was 0.68 sdu (though this is below the 0.96 sdu benchmark for comparing FonF/S instruction with a control with no focus on the target forms). An independent samples t test (table 5.16) also showed that class A's pre-post gains were statistically significantly higher than class C's¹⁴.

t	df ¹⁵	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
					Lower	Upper
2.575	56	.013	6.55490	2.54557	1.45552	11.65429

Table 5.15 Independent samples t test to compare the gains made by classes A and C between pre and post reading tests

As discussed in section 5.1.7, a background ethos of teach and test grammar at school 1 cannot, therefore, entirely explain the extent of the gains made by learners in class A. This implies that both the EI and PI learners must have benefited in some way from the experimental intervention.

5.2.7 Summary of 5.2: Analysis of the reading tests

The analyses above have suggested that in the whole dataset there were statistically significant differences between all pairs of tests. This was also the case for the intact classes A and C. Class had a statistically significant effect on the scores, and the findings for this were particularly conclusive between the post and dp tests. The

¹⁴ The reader is reminded that there were no significant differences between the EI and PI groups scores in class A.

¹⁵ Equal variances could be assumed.

effect of instructional group alone may have had a tendency to affect the learners' scores, with PI learners' scores being higher at post test and their gains probably being maintained at dp test. The analyses found that the learners' scores also depended on the class AND the instructional group they were in. EI learners in class B did not make any gains in their interpretation of written verb inflections, whereas the EI learners in class A made gains between all tests. The PI learners in classes A and B made statistically significant gains between pre and post test, and maintained these gains at dp test.

5.3 Summary of chapter 5 and discussion

This chapter has presented a range of analyses which tended to suggest that for the reading scores and, to a lesser extent, for the listening scores, PI had a statistically significant beneficial impact when compared to EI in class B. However, in class A, PI and EI learners made equivalent gains. As the EI learners in class A were able to make gains during the experimental intervention, this could suggest that they were detecting the target verb inflections, incidentally to the task that forced their attention on lexical items in the input. In terms of IP theory learners may have been attending to items of low Communicative Value (CV) whilst also attending to items of higher CV (e.g. lexical items) and/or overall sentential meaning. In contrast, in class B, EI learners seemed unable to detect verb inflections whilst also interpreting items of higher CV as they made no learning gains during the study. Equivalent learners, in the same class, receiving PI seemed to begin to process verb inflections, according to the measures used in this study, and they maintained their learning gains after approximately 14 weeks.

Clearly, these are contradictory findings with respect to the effect of EI in classes A and B, and some explanation is required. The analysis in section 4.2 showed that learners in class B (in school 2) were at an earlier developmental stage than the learners in school 1, in terms of interpretation and production of verb inflections (and possibly also in terms of metalinguistic knowledge and/or vocabulary range). IP theory would suggest that the higher developmental stage of EI learners in class A may have allowed attentional resources to be devoted to detecting items of low CV

(i.e. verb inflections) incidentally to interpreting lexical items and/or overall sentential meaning in the EI activities. It was also acknowledged that the EGI at the start of each unit of work, combined with the input flood in the EI activities, may have helped to activate their language knowledge / use, particularly given that these learners were accustomed to explicit grammar presentations and practice (see section 4.1). It was also suggested that a school / test effect in school 1 may also explain, to some extent, the improvement made by the EI group in class A. However, this cannot account entirely for their gains because the non-active control (class C), which had the same pre test scores and background as class A, made statistically significantly smaller gains than class A in both listening and reading measures.

Differences in general teaching and learning characteristics in the classes prior to the study may also have given learners in class A greater propensity than learners in class B to note (in a general sense and/or ‘detect’) verb inflections in the EI activities¹⁶. Section 4.1 described how learners in class A were familiar with planned sequences of grammar teaching activities which aimed to raise their awareness of a few specific features (and of verb inflections in particular) in the input and output. They were also accustomed to being asked about aspects of language beyond the minimum required by the task. As learners probably expected sequences of activities to be related in terms of language form, they may have tended to search for the features presented in the EGI in the EI tasks that followed it. The portraits of lessons in class B (outside the experimental intervention) suggested that learners in this class would not be as accustomed to such a sequencing of activities. Although explicit grammar explanations were frequent, they were unpredictable in terms of the activities preceding or following them and they could cover a wide range of features, probably due to their more ad hoc / reactive nature. In addition, not all learners appeared to attend to these explanations and comprehension of them often remained unchecked. As a result of some or all of the features summarised above, EI learners in class B may have been less likely to detect verb inflections in the EI materials¹⁷. It

¹⁶ The conditions *during* the intervention in both classes were similar (e.g. all materials were completed and learners in both classes appeared to be equally engaged in the tasks, see sections 4.1 and 4.3).

¹⁷ This scenario was offered some support by two of the EI learners in class B who commented that they didn’t understand why the activities after the EGIs were not related to the EGIs (suggesting that

is emphasised that these specific suggestions are tentative but they are offered to acknowledge the potential impact of more general teaching techniques on the results of this study.

the input flood of exemplars was not sufficient for them to practice the grammar point featured in the EGI).

Chapter 6 Results and Analysis of the Production Language Tests

Introduction

In this chapter quantitative analyses are used to assess the impact on the writing and speaking tests of the following independent variables: time of test (pre, post and delayed post test), type of instruction (PI, EI and class C), class (A, B or C), as laid out in research questions 2, 3 and 4. The analyses use very similar procedures to those in the previous chapter. In addition, this chapter also includes two other types of analysis.

First, some assessment of the validity and reliability of the test measurements was made by exploring whether the number of obligatory contexts (defined in section 3.6.6) produced during the oral and written narrative measurements had any relationship with class, instructional group or the actual writing score (i.e. target-likeness of inflection). These analyses were necessary in order to assess whether the class, instructional type or ability to inflect verbs correctly affected the number of verbs used by learners and, therefore, the calculation of their final production score.

Second, analysis was done to explore the speaking scores in terms of the type of discourse required (i.e. sentence or narrative level), as performance on oral narrative measures may have obtained lower scores than limited response sentence level tasks, as found in VanPatten & Sanz (1995)¹. This analysis was additionally relevant for this study to investigate whether the speaking scores were influenced by their composition, as some learners' scores consisted of sentence level tasks only and others of sentence and narrative level tasks combined (see sections 3.6 and 6.2 for fuller descriptions and rationales of how the scores were obtained).

¹ Other PI studies have not used oral narrative measures.

6.1 Analysis of the writing tests

The total writing scores were turned into a percentage. The total raw score possible in the sentence level tasks was 56 and this was added to the number of obligatory contexts (briefly, verbs referring to a subject) each learner created for themselves in the narrative level tasks². The definition of obligatory contexts and how the writing was scored are outlined in section 3.6.6 and in appendix 23.

The reader is reminded that during the written post tests, class B were visibly and audibly distressed at having to do a French test (see section 3.5.1).

6.1.1 *Can parametric tests be used?: normality of the distribution of the data*

Several samples showed a statistically significant difference to the normal distribution at the 95% confidence level (see appendix 29). However, at the 99% confidence level³ only the merged EI group at pre and dp test and the PI group in class B at pre test were not normally distributed. Parametric tests were therefore carried out, and non-parametric tests were performed to supplement these, in line with analyses in chapter 5, and as explained in section 3.8.2.

6.1.2 *Analysis of issues surrounding the reliability and validity of the writing tests*

Inferential statistics were used to compare the number of obligatory contexts produced at pre, post and dp test in the various sample groups in the study⁴. It was found that the test variable did impact on the number of obligatory contexts produced i.e. more were produced at post test, and this did not change at dp test. In most cases, the class and merged instructional grouping variables did not have a statistically significant impact on the production of obligatory contexts. This implies therefore that, in general, it is unlikely that any particular samples of the scores were based on

² A small number of learners did not respond to an occasional sentence item (probably because they did not know the lexical verb). Analyses were carried out both including and excluding this data and, although the raw scores were very slightly different, no differences were found in any of the statistical tests done. In the presentation here, such instances have been included.

³ i.e. increasing the chances of accepting the null hypothesis of no difference between the sample and a normal distribution

⁴ Due to space limitations, appendix 35 contains the results of these analyses.

a very small number of contexts and others on significantly larger samples of language. However, it was found that in class B at post test, the PI group produced more obligatory contexts than the EI group and this difference was borderline at dp test. This appears to be in line with the actual scores (chapter 5 and the remainder of this chapter), in that in class B PI generally made a positive impact on the *accuracy* of interpretation and production of verb inflections at post test (and usually also at dp test).

These findings are also relevant as, as mentioned in section 4.3, teachers A and C suggested that EI may be more suitable for teaching vocabulary and PI for teaching grammar. If this were the case, it might have been expected that EI learners would produce more obligatory contexts than PI learners. However, this was not found to be the case, and in fact, the PI learners in class B were more likely to produce verbs with a subject than EI learners at post test. Nevertheless, it is acknowledged that a precise measure of vocabulary range was not taken and this would be required before more definitive claims can be made regarding the relative impact of each input approach on lexical acquisition.

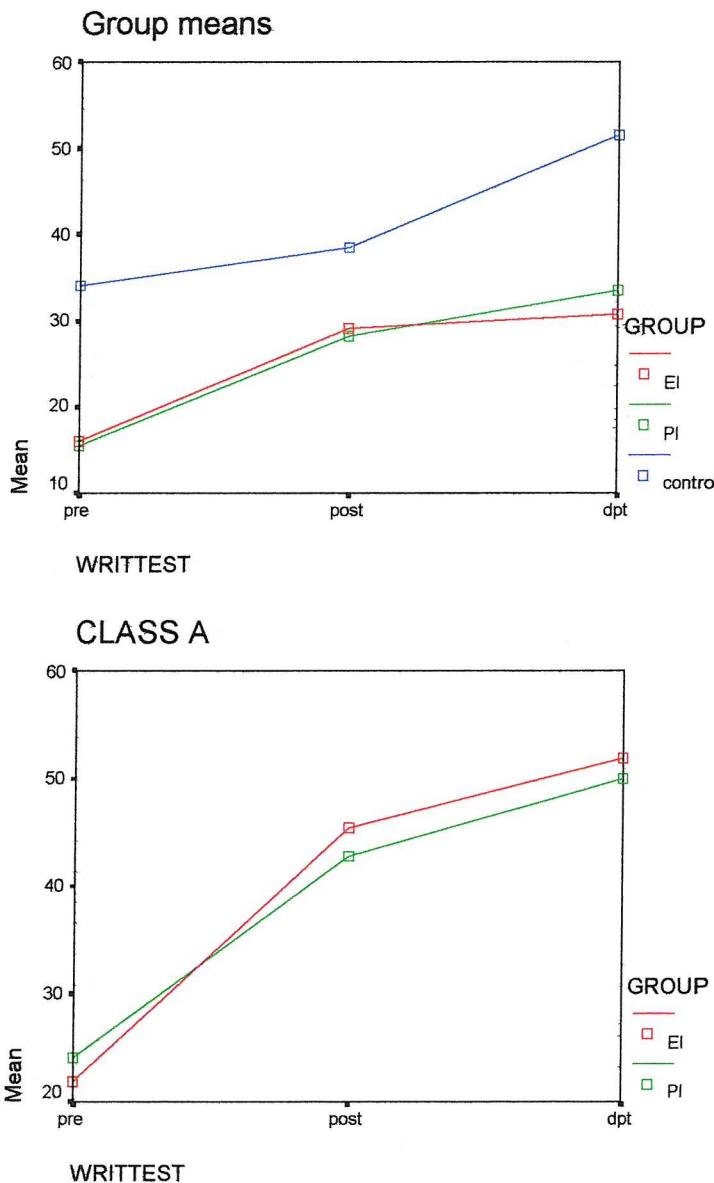
Correlations between the number of contexts generated and the scores were also calculated. If the number of contexts produced correlates positively with the target-likeness of the pupils' productions (suggesting some relationship between the quantity and quality of productions) then claims regarding the nature of the influence of instructional type on the scores would be affected. Little evidence was found to suggest any correlation between the number of contexts produced and the accuracy of the verb inflections. This suggests that the tests, in general, elicited obligatory contexts from the pupils regardless of their developmental stage in terms of verb inflection i.e. learners were willing to 'have a go' at the tests. The only exception to this was a positive correlation found in the PI group in class B at dp test (Pearson's $r = 0.557$, p (two tailed) = 0.039, $n=14$). As suggested above, this correlation may indicate some kind of impact of PI in terms of encouraging learners in class B to produce more verbs than EI learners, *and* that these verbs are more likely to have target-like inflections. However, such suggestions are tentative at this stage as the n

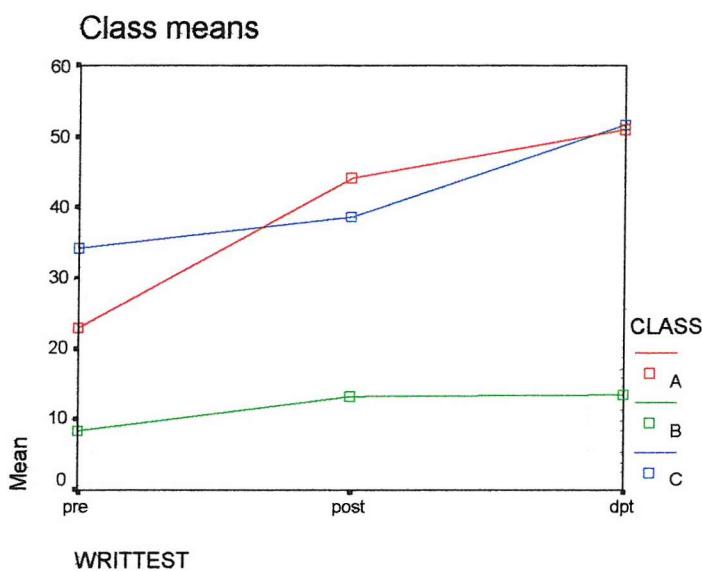
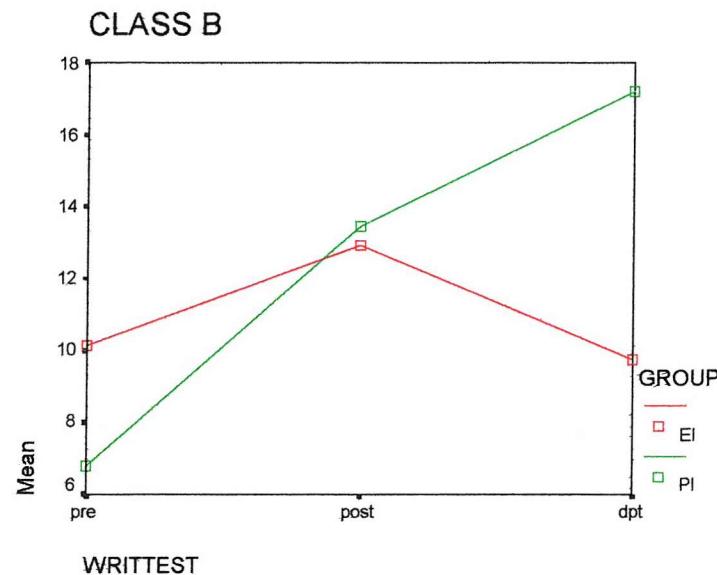
was small and the effect was not found at post test (although this may have been affected by class B's lack of motivation during the written post tests).

6.1.3 Descriptive statistics

Graphical presentation of the results and effect sizes are given in this section (see appendix 31 for the means, standard deviation and n).

a) Graphs





b) Effect sizes on gains between tests

The measure of effect size used here (*Cohen's d*) was described in section 3.8.6. As in chapter 5, a benchmark of 0.75 standard deviation units (found in Norris & Ortega 2000 for effect sizes where the comparison group interacted with materials in which the target forms were embedded) will be used here to help interpret the effect sizes found. In order to eliminate the differences between the groups at pre test, the effect sizes of the gains between the tests were calculated (table 6.1).

<i>Effect size of...</i>	<i>Pre- post gains</i>	<i>Post- dp gains</i>	<i>Pre - dp gains</i>
merged PI over merged EI	-0.07	0.16	0.12
class A PI over EI	-0.28	0.06	-0.22
class B PI over EI	0.41	0.27	0.98

Table 6.1 *Effect sizes for gains in writing scores*

There was little difference between the effect sizes of the gains made by the merged instructional groups. Similarly, in class A neither instructional group had a clear advantage in terms of the effect size of their gains.

In class B however, by the time of dp test, the PI gains had an effect size larger than the 0.75 sdu benchmark and of the magnitude found by Norris & Ortega (2000) when the comparison group had *no* exposure to the target feature (0.96 sdu).

6.1.4 Repeated measures ANOVA to assess the impact of test, class and instructional group on the writing scores

A repeated measures ANOVA was carried out, with the three writing tests (WRIT) as repeated measures and with class and instructional group (GROUP) as between-subjects factors. The assumption of equality of variances was not met (appendix 32) and some of the data samples did not have a normal distribution,. The analysis here can therefore only be used cautiously and non-parametric tests are presented later.

Source	Type III Sum of Squares	df	Mean Square	F	Sig. ⁵
WRIT	11363.490	2	5681.745	55.931	.000
WRIT * GROUP	119.183	2	59.592	.587	.557
WRIT * CLASS	3822.932	2	1911.466	18.816	.000
WRIT * GROUP * CLASS	385.024	2	192.512	1.895	.154

Table 6.2 *Repeated measures ANOVA for the writing scores*

⁵ Mauchly's test of sphericity showed that the variances of the differences between conditions are not statistically significantly different (Mauchly's W 0.924, approx Chi Square, 5.834, df 2, p=0.054), and so no correction to p was required.

This ANOVA shows that there were statistically significant differences between the tests and that the impact of class on the writing test scores (the interaction writ*class) was statistically significant. Simple and repeated planned contrasts were carried out to explore the source of these differences (see table 6.3).

Source	WRITTEST	Type III Sum of Squares	df	Mean Square	F	Sig.
WRIT	Post vs. Pre	8614.596	1	8614.596	37.845	.000
	Dp vs. Pre	22286.726	1	22286.726	95.034	.000
	Post vs. Dp	3189.149	1	3189.149	21.640	.000
WRIT * GROUP	Post vs. Pre	3.238	1	3.238	.014	.905
	Dp vs. Pre	153.260	1	153.260	.654	.421
	Post vs. Dp	201.051	1	201.051	1.364	.246
WRIT * CLASS	Post vs. Pre	3679.154	1	3679.154	16.163	.000
	Dp vs. Pre	7203.236	1	7203.236	30.716	.000
	Post vs. Dp	586.407	1	586.407	3.979	.050
WRIT* GROUP * CLASS	Post vs. Pre	259.486	1	259.486	1.140	.289
	Dp vs. Pre	763.012	1	763.012	3.254	.075
	Post vs. Dp	132.574	1	132.574	.900	.346

Table 6.3 Simple and repeated contrasts of the test variable for the writing scores

The planned contrasts suggest that there were statistically significant differences between all pairs of the test. The interaction writ*class is statistically significant between pre and post and between pre and dp test. The differences between the post and dp test in the different classes were only of borderline significance (p=0.05). The interaction writtest*group is non-significant between all pairs of tests.

As with the interpretation measures analysed in chapter 5, class C was excluded in order to assess the impact of this class on the findings above i.e. did excluding class C alter the effect of class, group or the interaction group*class on the test scores? As with the other measures, this analysis confirmed that class C had not affected the findings of the above analysis (see appendix 33 for details).

As there were differences between some of the samples at pre test (some of which were statistically significant and others which were not) an analysis of the gains

between the tests was conducted which also supported the findings above (appendix 34).

6.1.5 Transforming the scores

When the scores are transformed by using the square root of the raw score, nearly all of the samples are normally distributed (see appendix 29)⁶. Also, the transformation of the scores resulted in more equal variances - the pre and post test scores could be considered equal at 99% confidence level (see appendix 32). Further transformation of the data is considered beyond the scope of this thesis mainly due to the fact that transformed data is difficult to interpret. This ANOVA is therefore presented as an indication of trends only.

Source	Type III Sum of Squares	df	Mean Square	F	Sig. ⁷
$\sqrt{\text{WRIT}}$	105.555	1.656	63.734	45.910	.000
$\sqrt{\text{WRIT}} * \text{GROUP}$	4.186	1.656	2.528	1.821	.172
$\sqrt{\text{WRIT}} * \text{CLASS}$	20.146	1.656	12.164	8.762	.001
$\sqrt{\text{WRIT}} * \text{GROUP} * \text{CLASS}$	8.627	1.656	5.209	3.752	.034

Table 6.4 Repeated measures ANOVA on the square root of the writing scores

This analysis suggests that the interaction between $\sqrt{\text{writing test}} * \text{group} * \text{class}$ is significant supporting the suggestion from the graphs above that the type of input instruction had a differential effect depending on the class learners were in. (The repeated measures ANOVA with non-transformed scores presented above produced a non-significant interaction.) The planned contrasts (table 6.5) suggest that the source of this difference is between the pre and dp test measures. The graphs suggest that learners in class B performed differently depending on which instructional group they were in between the pre and dp tests, whereas the other groups and classes do not have distinct patterns.

⁶ The only exception was the overall post test results (though at the 99% confidence level these could be considered normally distributed). In any case, this gross sample does not affect any of the conclusions drawn.

⁷ Mauchly's test of sphericity suggested that sphericity does not hold (0.792, 17.219, 2, $p < 0.05$) and so the Greenhouse-Geisser estimate of sphericity (0.828) is used in order to calculate the df which is used to find the p value for the F statistic (in any case, all the estimates of sphericity gave the same results in terms of statistical significance at the 95% confidence level).

Source	\sqrt{WRIT}	Type III	df	Mean Square	F	Sig.
		Sum of Squares				
\sqrt{WRIT}	Post vs. Pre	94.630	1	94.630	34.280	.000
	Dp vs. Pre	201.940	1	201.940	70.044	.000
	Post vs. Dp	20.095	1	20.095	16.025	.000
$\sqrt{WRIT} * GROUP$	Post vs. Pre	.735	1	.735	.266	.607
	Dp vs. Pre	7.964	1	7.964	2.762	.101
	Post vs. Dp	3.860	1	3.860	3.078	.083
$\sqrt{WRIT} * CLASS$	Post vs. Pre	17.215	1	17.215	6.236	.015
	Dp vs. Pre	38.872	1	38.872	13.483	.000
	Post vs. Dp	4.350	1	4.350	3.469	.066
$\sqrt{WRIT} * GROUP * CLASS$	Post vs. Pre	5.363	1	5.363	1.943	.167
	Dp vs. Pre	17.176	1	17.176	5.957	.017
	Post vs. Dp	3.343	1	3.343	2.666	.107

Table 6.5 Simple and repeated contrasts of the test variable for the square root of the writing score

Carrying out the above analyses without class C did not make any substantial difference to the results.

6.1.6 Analysing classes separately

Although the analyses above have suggested that there may be a tendency for PI to have a differential impact depending on the class learners were in, this analysis could not be considered conclusive as some interactions showed borderline statistical significance and an analysis of the use of transformed data found slightly different results. In addition, as the analysis in section 4.2 suggested that class B was from a statistically significantly different population at the outset, separate repeated measures ANOVAs were carried out on classes A, B and C.

Levene's test of equality of error variances showed this assumption was upheld in all tests in classes A and in the pretest in class B. In post and dp tests in class B the variances were equal at the 99% confidence level. The repeated measures ANOVA can be used as an indication of trends.

The ANOVA found significant differences between the tests. The interaction of writ*group was not significant in class A (planned contrasts are therefore superfluous

for this interaction). In class B, the repeated measures ANOVA suggests that both the writing test and the interaction writ*group are statistically significant.

	Source	Type III Sum of Squares	df	Mean Square	F	Sig. ⁸
Class A	WRIT	12321.788	1.645	7491.020	44.588	.000
	WRIT * GROUP	98.991	1.645	60.181	.358	.659
Class B	WRIT	409.502	2	204.751	4.091	.023
	WRIT * GROUP	388.727	2	194.364	3.884	.027
Class C	WRIT	4074.494	2	2037.247	18.196	.000

Table 6.6 Repeated measures ANOVA for class A, B & C's writing test scores

	Source	WRITTEST	Type III Sum of Squares	df	Mean Square	F	Sig.
Class A	WRIT	Post vs. Pre	12945.640	1	12945.640	41.968	.000
		Dp vs. Pre	22667.159	1	22667.159	61.603	.000
		Post vs. Dp	1352.566	1	1352.566	8.862	.006
Class B	WRIT	Post vs. Pre	577.861	1	577.861	6.341	.019
		Dp vs. Pre	648.602	1	648.602	5.125	.033
		Post vs. Dp	2.042	1	2.042	.025	.876
	WRIT * GROUP	Post vs. Pre	96.863	1	96.863	1.063	.313
		Dp vs. Pre	757.017	1	757.017	5.982	.022
		Post vs. Dp	312.301	1	312.301	3.781	.064
Class C	WRIT	Post vs. Pre	473.502	1	473.502	1.733	.200
		Dp vs. Pre	7525.979	1	7525.979	39.126	.000
		Post vs. Dp	4224.000	1	4224.000	20.481	.000

Table 6.7 Simple and repeated contrasts for class A, B and C's writing test scores

Planned contrasts suggested there were improvements between all tests in class A.

In class B, planned contrasts found that there were statistically significant differences between the pre and post tests and these were maintained at dp test. However, there were no differences between post and dp tests. They also found that the effect instructional group had on test was significant between pre and dp tests, and approaching significance between post and dp tests. Again, it is suggested that a reason for this may be the distress during the written post tests in class B, adversely affecting performance, but that at dp test, the PI learners were able to improve further

⁸ Mauchly's test of sphericity showed that the data from class A did not uphold this assumption (0.784, 6.324, 2, p=0.042). The Greenhouse-Geisser correction (0.822) was used to calculate p. This was not necessary for the other classes.

on the gains made at post test in order to make a statistically significant difference compared to the pre test scores. The EI learners in class B made no statistically significant gains in the target features throughout the study.

In class C there were no statistically significant gains between pre and post test in this measure. This is the only measure where this is the case. However, between post and dp tests statistically significant gains were made.

6.1.7 Non-parametric tests

The non-parametric tests supported the findings above (see appendix 27). The tests found that in the whole dataset there were significant differences between each test. In the intact class A there were also statistically significant differences between each test. In class C, there were statistically significant differences between the post and dp tests and between the pre and dp tests, though not between the pre and post test. The Friedman test showed that for the intact class B, the differences between the tests were approaching significance ($p=0.054$) and the paired Wilcoxon tests showed that there were significant gains between pre and post and between pre and dp test, though not between the post and dp test. Both the merged EI and PI groups made gains between the pre and post and between the pre and dp test. Only the merged PI group had significant differences between the post and dp test. In class A, both PI and EI groups made pre - post and pre – dp test gains. The post – dp test differences were approaching significance (and one tailed tests⁹ showed statistical significance). In class B, the EI group made no gains between any of the tests. The PI group made gains between the pre and post and maintained these gains at dp test. There was no statistically significant difference between the post and dp test scores.

6.1.8 Discussion and analysis of class C, the non-active control

As in sections 5.1.7 and 5.2.6, the gains made by classes A and C between pre and post tests were compared to explore the extent to which a school effect can explain the gains made by the EI (and possibly PI) learners in class A. The repeated measures ANOVA and planned contrasts on class C (tables 6.6 and 6.7 above) showed that the gains made by class C between pre and post tests were not

statistically significant. Furthermore, the effect size of the pre-post gains made by class A compared to class C was 0.99 sdu, just above the benchmark 0.96 sdu (appropriate because class C did not have FonF/S between the pre and post tests). An independent samples t test (table 6.8) also showed that class A made significantly greater gains than class C. Therefore, a school effect cannot explain the PI and EI learners' gains in class A between pre and post tests. As in chapter 5, it is suggested that it is likely that the EI learners in class A were able to benefit from the input flood (and possibly the EGI) in the EI, in ways that EI learners in class B were not.

t	df ¹⁰	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
					Lower	Upper
3.709	54	.000	16.58169	4.47117	7.61754	25.54583

Table 6.8 *Independent samples t test to compare gains made by classes A and C between pre and post writing tests*

6.1.9 Summary of section 6.1: Analysis of the writing tests

This section has presented a range of analyses which have shown that instruction, regardless of class, did not have an overall effect on the writing scores. However, learners' scores depended on the class and the instructional group they were in. In class A, instructional type had no differential impact on gains made by learners – both EI and PI learners made statistically significant gains. In class B only the PI learners made statistically significant gains between pre and post and maintained these gains at dp test. Learners in class C did not make statistically significant gains between pre and post test (though they did between post and dp test, probably due to the FonF/S instruction on verb inflections in the perfect tense). Therefore, a school effect cannot be used to explain the gains made by the PI and EI groups in class A between pre and post tests: it is likely that they both benefited from their input-based instruction.

⁹ testing the hypothesis that dp test scores would be higher than post test scores

¹⁰ Equal variances could be assumed

6.2 Analysis of the speaking tests

As described in chapter 3, 16 learners from each of classes A and C and 17 from class B undertook the sentence level speaking tasks (with almost equal numbers of PI and EI learners¹¹). These sentence level tasks carried 54 available points (27 items each with 2 points available). Scores were turned into a percentage using 54 as the denominator.

Of these 16 learners from each class, 8 also undertook narrative and guided conversation tasks (henceforth referred to as narrative tasks), where scoring was dependent, to some extent, on how many obligatory contexts each learner produced. For these learners, their sentence scores were added to their narrative task scores, and an 'individual' denominator was calculated for each learner by adding the 54 points available for the sentence items to the number of obligatory contexts created by the learner ($x2$, as two points were available for each obligatory context). See appendix 24 for details of how the written interlanguage was scored.

Therefore, half the oral production scores are derived from the sentence level tasks only and the other half consist of the sentence level tasks combined with the scores from the narrative tasks. Both were calculated as a percentage.

6.2.1 Normality of the data: Can parametric tests be used?

Several samples of data from the speaking tests could not be considered as being normally distributed, as shown by the results of the K-S test in appendix 29. The pre, post and dp test scores were heavily positively skewed (many more learners obtaining low scores than would be expected in a normal distribution). The samples that cannot even be considered as normally distributed at the 99% level are: the overall pre, post and dp tests, Class A pre test, merged PI group pre and dp tests, and Class A PI group pre and post tests. This, along with the small sample sizes, indicated that the results from parametric tests should be interpreted with caution and that non-parametric tests were required to supplement them.

¹¹ see section 3.6.4: there was one more EI learner than PI learner in class B due to an administrative error.

6.2.2 Analysis of issues surrounding the reliability and validity of the speaking tests

The mean score available in the speaking tasks was 117.3. To give an approximate picture of the composition of the total speaking score, table 6.9 illustrates the total scores available for each speaking task (the *mean* numbers of contexts produced for the narrative tasks are given).

Sentence level			Narrative / conversation					
Present tense, regular, er	Present tense, irregular, 3 rd person	Perfect tense, regular, avoir & être	Present tense (std. dev.)			Perfect tense (std. dev.)		
			pre	post	dp	pre	post	dp
16	16	22	27.1 (6.5)	31.7 (5.1)	30.7 (7.4)	32.3 (5.9)	34.3 (4.4)	34.1 (5.4)
overall mean = 29.8						overall mean = 33.5		
54			combined mean = 63.3					

Table 6.9 Total scores available for oral production tasks

a) Production of obligatory contexts in the oral narratives/guided conversation

As with the written narrative measures, it was ascertained whether the test instrumentation was consistent in terms of the elicitation of obligatory contexts across the different tests, classes and instructional groups. The statistical details of this analysis are given in appendix 36. It was found that the number of obligatory contexts produced increased at post test and this was maintained at dp test. The class and group learners were in did not impact on the number of contexts produced. This implies that the oral narrative tests were robust in their elicitation of obligatory contexts across different samples of learners. It also eliminates the possibility that some of the scores were based on a very small number of contexts and others on significantly larger samples of language. This analysis also suggests that EI learners did not have a disadvantage in terms of their attempts at producing verbs with a subject, addressing, to some extent, teachers A and B's suggestion that EI may be more effective for the teaching of lexical items.

In order to check whether there was any correlation between the number of contexts produced and the accuracy of production, further analysis was necessary, using Spearman's Rho (a non parametric test), as the n was small and some of the data sets were not normally distributed (see appendix 36 for details of this analysis). It was found that, overall, at pre test learners produced obligatory contexts without necessarily producing accurate inflections (i.e. there was no statistically significant correlation). At post and dp tests, overall, learners were more likely to produce a correct inflection the more contexts they produced. At post test it was likely for this to be the case in the merged PI group and in class A. However, the statistical significance of this tendency in the merged PI group at post test is likely to be heavily influenced by the fact that half these learners were in class A (where the relationship was clear), as neither instructional group in class B showed this tendency. Furthermore, the tendency was apparently reversed at dp test (i.e. the merged EI group was more likely to produce accurate inflections the more obligatory contexts they produced, and the tendency was no longer significant in class A). It is therefore argued that there is no convincing evidence, especially given the small sample size, of a pattern regarding the effect of instructional type and/or class on the learners' production of contexts and their ability to inflect verbs accurately. In summary, for the purposes of this study the oral production measure can be said to be sufficiently robust in terms of eliciting reliable samples across different instructional groups and classes.

b) Exploring the composition of the speaking score: sentence level versus narratives?

There is some debate that sentence and narrative level oral tasks elicit different types of learner language, particularly in terms of its accuracy (e.g. Truscott 1998, Norris & Ortega 2000, VanPatten & Sanz 1995). The oral measure used in this study consisted of a combination of sentence level and narrative / guided conversation tasks. It was of interest to assess whether these task types obtained similar results, both to inform this debate and to assess the internal validity of the oral score. Sentence level and narrative scores from learners who completed both types of tasks were analysed using Pearson's product moment correlation (see table 6.10).

	Pearson's correlation of sentence score with narrative score	Sig. (2-tailed)	N
Pre	.748	.000	23
Post	.586	.003	23
Dp	.598	.003	23

Table 6.10 Correlations between the sentence and narrative oral scores

Table 6.10 shows a high degree of positive correlation between the two measures.

Paired samples t tests¹² were also carried out in order to see if the two sets of measures had produced scores which came from statistically significantly different populations. It was found that two scores were consistently statistically similar (see table 6.11).

Test	Mean	Std. Deviation	Std. Error	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Mean	Lower	Upper		
Pre	-1.0571	10.03645	2.09274	-5.3971	3.2830	-.505	22	.619
Post	-2.5521	17.20833	3.58818	-9.9935	4.8894	-.711	22	.484
Dp	5.0045	17.84722	3.72140	-2.7132	12.7222	1.345	22	.192

Table 6.11 Paired samples t tests to compare the sentence and narrative oral scores

This finding contradicts the finding in VanPatten & Sanz 1995 that learners were more likely to produce accurate language in sentence level tasks than in the narratives. It is argued that in this study the two ways of arriving at a speaking score did not obtain statistically different results and could be combined to produce one valid measurement, which was possibly more robust as it comprised two types of performance that are often considered to be different.

c) *Checking whether the sample of learners were similar*

In order to confirm that learners who undertook both the sentence and narrative level tasks and those who undertook just the sentence level tasks were from the same population independent sample t tests were carried out¹³.

¹² As some of the data was not normal, non-parametric tests (Wilcoxon matched pairs signed ranks test) were also carried out. The same results were found.

¹³ To select pupils who would do the narrative tasks, 2 pupils who had undertaken the oral sentence level tasks were chosen from each quartile of the reading and listening scores, see section 3.6.4.

	t	df ¹⁴	Sig. (2- tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						Lower	Upper
Pre	.187	46	.853	.8164	4.37022	-7.98038	9.61323
Post	-.194	46	.847	-1.0692	5.50932	-12.15893	10.02045
Dp	-1.027	45	.310	-6.3389	6.17408	-18.77414	6.09634

Table 6.12 Independent sample t tests to compare the oral sentence level scores of learners who took only the sentence level tasks and learners who took both the sentence and narrative level tasks

It was found that the sentence level scores of learners who had taken both the sentence tasks and the narrative tasks can be considered the same as the scores from learners who only took the sentence level tasks, indicating that the two samples of learners were from the same population¹⁵.

This section has shown that although the spoken production score was derived from two types of activity for some learners but from one for other learners, it can be considered as a valid measurement at pre, post and dp tests. This suggests that if any impact of instruction on oral performance is found, then narrative level performance may have been affected in similar ways to sentence level performance. However, a specific investigation of the impact of instructional type on sentence versus narrative level oral performance is beyond the scope of this thesis. As emphasised in chapter 3, using both types of task for the calculation of an oral measurement is intended to increase the sample of language and potentially different performance types elicited.

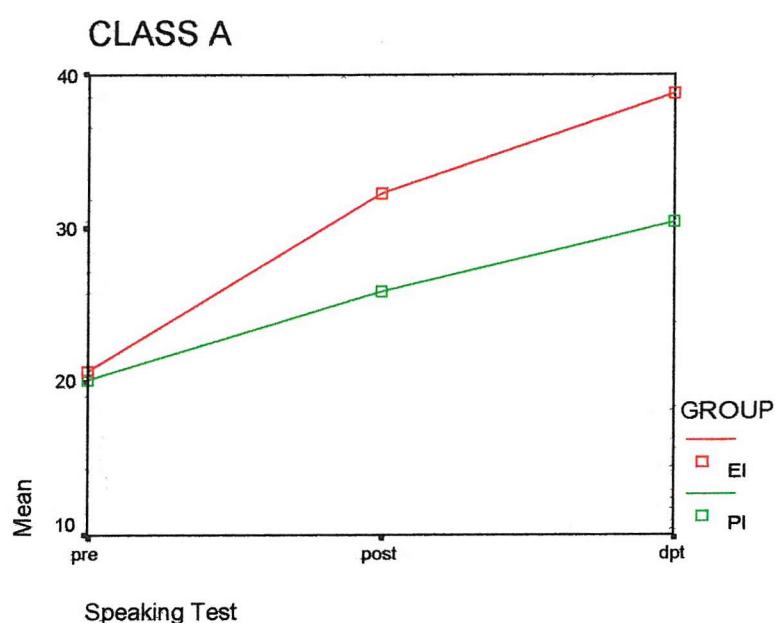
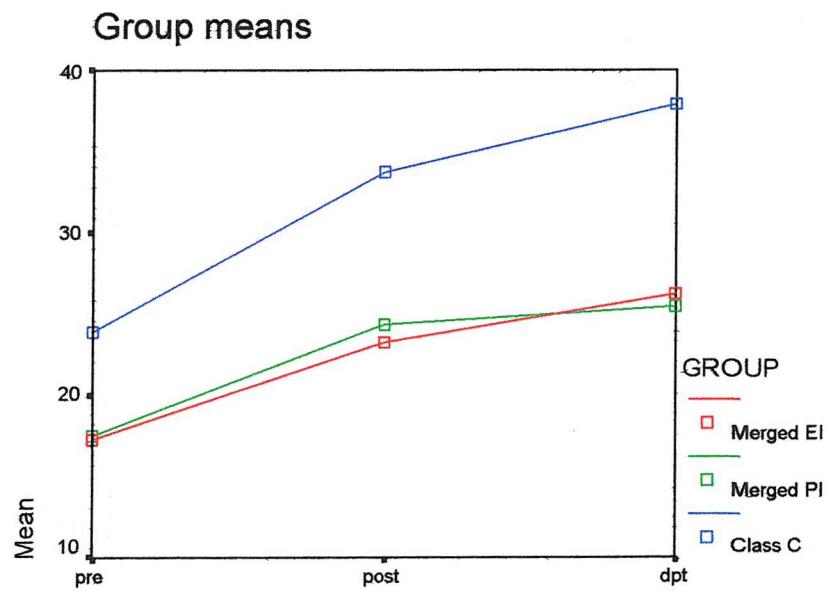
The following section explores the speaking scores in terms of the impact of different instructional types and classes.

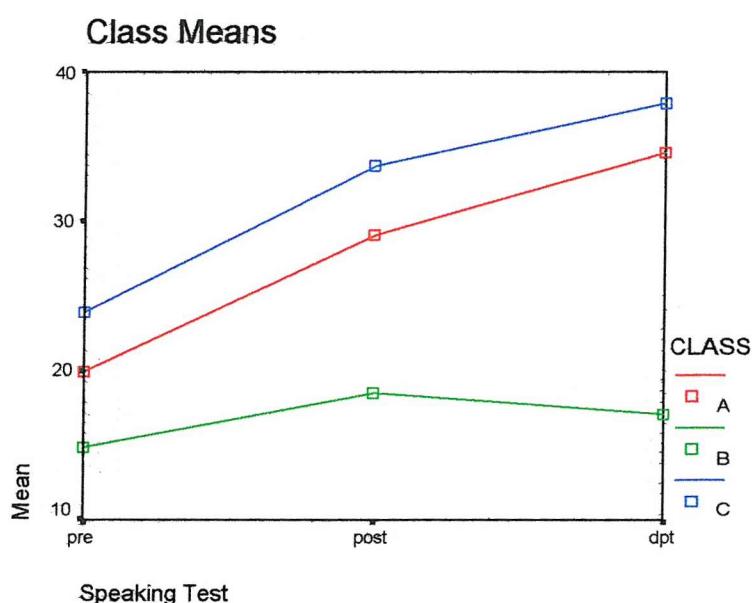
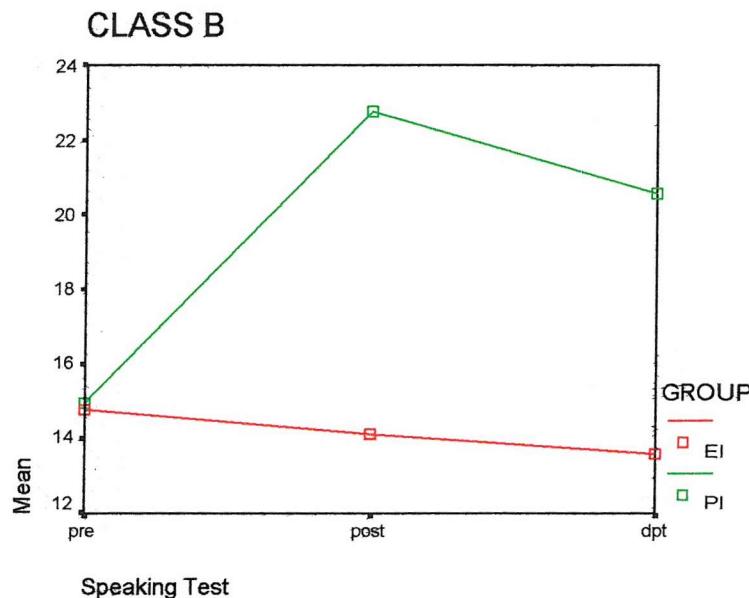
¹⁴ Equal variances could be assumed

¹⁵ Non-parametric tests (Mann-Whitney U test for independent samples) were carried out in order to confirm these findings as some of the data was not normally distributed. The same results were found and so are not presented here

6.2.3 Descriptive statistics

a) Graphs





b) Effect sizes for gains between tests

<i>Effect size of...</i>	<i>Pre- post gains</i>	<i>Post- dpt gains</i>	<i>Pre - dpt gains</i>
merged PI over merged EI	0.05	-0.22	-0.15
class A PI over EI	-0.69	-0.16	-0.62
class B PI over EI	0.87	-0.19	0.98

Table 6.13. *Effect sizes of gains between speaking tests*

It can be seen that there is little substantial effect size for either instructional type in the gains made in the merged EI and PI groups.

However, as with previous analyses, the effect size for PI is different in the two classes. In class B by the time of the dp test, PI had an effect size that was considerably higher than the appropriate benchmark (0.75 sdu) and even of the magnitude of the benchmark for comparing FonF/S to controls with no exposure to the target forms (0.96). In contrast, in class A EI had an effect size over PI. This contradicts the finding in Norris & Ortega (2000) that in general FonF/S interventions had positive effect sizes of a mean of 0.75 sdu over interventions such as the EI in this study.

6.2.4 Repeated measures ANOVA to assess the impact of test, instructional group and class on the speaking scores

Levene's test of equality of error variances (appendix 32) suggests that only the dp test scores did not have equal variances. The ANOVA can therefore provide an indication of trends.

Source	Type III Sum of Squares	df	Mean Square	F	Sig. ¹⁶
SPK	2577.033	2	1288.516	22.895	.000
SPK * CLASS	630.301	2	315.151	5.600	.005
SPK * GROUP	14.065	2	7.032	.125	.883
SPK * CLASS * GROUP	319.465	2	159.733	2.838	.064

Table 6.14 Repeated measures ANOVA for speaking test scores

The repeated measures ANOVA suggests that there are statistically significant differences between the different overall tests and between the different tests depending on the class the learners were in. There was no overall statistically significant difference depending only on the instructional group the learners were in. However, the interaction between speaking test*class*group is approaching statistical significance and suggests that the score learners achieved at different tests

depended on both the class they were in and the input group they were assigned to.

Planned contrasts of these effects produced the results presented in table 6.15.

Source	SPKTEST	Type III Sum of Squares	df	Mean Square	F	Sig.
SPK	Post vs. Pre	2582.093	1	2582.093	27.280	.000
	Dp vs. Pre	4806.274	1	4806.274	42.233	.000
	Post vs. Dp	342.731	1	342.731	2.652	.111
SPK * CLASS	Post vs. Pre	249.190	1	249.190	2.633	.112
	Dp vs. Pre	1255.627	1	1255.627	11.033	.002
	Post vs. Dp	386.086	1	386.086	2.988	.091
SPK * GROUP	Post vs. Pre	6.400	1	6.400	.068	.796
	Dp vs. Pre	7.684	1	7.684	.068	.796
	Post vs. Dp	28.109	1	28.109	.218	.643
SPK * CLASS * GROUP	Post vs. Pre	468.626	1	468.626	4.951	.031
	Dp vs. Pre	489.542	1	489.542	4.302	.044
	Post vs. Dp	.228	1	.228	.002	.967

Table 6.15 Simple and repeated planned contrasts between speaking tests

This suggests that the interaction between speaking test*class*group was statistically significant between pre and post and between pre and dp tests, yet not between post and dp tests. This suggests that the combined impact of instructional group and class maintained its impact on speaking scores at dp test.

The class effect alone was not statistically significant between pre and post test, and between post and dp test but was significant between pre and dp test - suggesting that over the course of the whole study, the class learners were influenced the scores.

In order to explore whether the results above were due to the effect of class C acting as a 'group' and influencing the interaction effects, the same procedures as above were carried out without class C. Again, the results suggested that class C did not affect the findings in terms of statistical significance and so are not presented here¹⁷ (see appendix 33).

¹⁶ Mauchly's Test of Sphericity showed that this assumption was upheld and therefore that no correction to the p value was required (Mauchly's W 0.968, approx Chi-square 1.347, 2, p=0.510).

¹⁷ The only difference was that according to the planned contrasts carried out without class C the interaction spktest*class*group between pre and dp test was only approaching significance at 95%

An analysis using the gains between the tests was carried out, in line with other measures, even though the differences between the samples at pre test were not statistically significant (oral production was the only measure where classes A and B had statistically similar pre test scores, see sections 3.3 and 4.2). The results supported the findings presented above (see appendix 34).

Transforming the scores using the square root of the speaking scores achieved a more normally distributed set of data (as shown in appendix 29), and the variances in the dp test variable were equal at the 99% confidence level. The only difference (in terms of statistically significant findings) between the analyses of the transformed and non-transformed scores was that in the former the interaction $\sqrt{\text{spktest}} * \text{class} * \text{group}$ was only approaching significance at the 95% confidence level ($F 3.005, 2, p=0.055$). However, the results have more validity as more of the necessary assumptions for carrying out this test were upheld, thus reinforcing the claim that learners' scores probably depended on their class and the instructional type they had.

6.2.5 Analysing classes separately

Although the analysis in section 4.2 did not show conclusively that there were significant differences between the classes in the speaking pretest, it did suggest that, in line with other measures, class B may have been from a different population. Separate repeated measures ANOVAs and planned contrasts (tables 6.16 and 6.17) were therefore carried out on the classes, as with the other measures.

Levene's test of equality of error variances showed this assumption was upheld in all variables (relevant for the ANOVAs in class A and B, appendix 32).

confidence level ($p=0.056$). However, this still suggests that there was a tendency for the type of instruction to have a differential impact at dp test depending on the class learners were in.

	Source	Type III Sum of Squares	df	Mean Square	F	Sig. ¹⁸
Class A	SPK	1748.914	2	874.457	11.682	.000
	SPK * GROUP	167.373	2	83.687	1.118	.341
Class B	SPK	107.760	2	53.880	1.413	.259
	SPK * GROUP	166.139	2	83.069	2.179	.131
Class C	SPK	1559.885	2	779.942	13.646	.000

Table 6.16 Repeated measures ANOVAs carried out on classes separately for the speaking tests

	Source	SPKTEST	Type III Sum of Squares	df	Mean Square	F	Sig.
Class A	SPK	Post vs. Pre	1333.130	1	1333.130	14.165	.002
		Dp vs. Pre	3427.986	1	3427.986	16.958	.001
		Post vs. Dp	485.627	1	485.627	3.177	.096
	SPK * GROUP	Post vs. Pre	180.138	1	180.138	1.914	.188
		Dp vs. Pre	305.518	1	305.518	1.511	.239
		Post vs. Dp	16.464	1	16.464	.108	.748
	SPK	Post vs. Pre	211.910	1	211.910	2.200	.159
		Dp vs. Pre	79.642	1	79.642	1.605	.225
		Post vs. Dp	31.730	1	31.730	.383	.545
	SPK * GROUP	Post vs. Pre	296.575	1	296.575	3.079	.100
		Dp vs. Pre	190.034	1	190.034	3.829	.069
		Post vs. Dp	11.807	1	11.807	.143	.711
Class C	SPK	Post vs. Pre	1454.569	1	1454.569	15.571	.001
		Dp vs. Pre	2960.360	1	2960.360	31.420	.000
		Post vs. Dp	264.725	1	264.725	1.705	.213

Table 6.17 Simple and repeated planned contrasts after repeated measures ANOVAs on separate classes for the speaking scores

The repeated measures ANOVAs and planned contrasts (table 6.16) showed that in class A 'test' had a significant effect on the scores between pre and post and between pre and dp tests but not between post and dp tests. This is perhaps surprising given the production practice of the perfect tense these learners had between post and dp tests. The instructional group had no impact on the scores in class A.

In Class B the ANOVA suggests that neither the test nor the interaction spk*group was statistically significant. Planned contrasts were, nevertheless, carried out to look

¹⁸ Mauchly's test of sphericity showed this assumption was upheld (class A: .826, 2.486, 2, p=.289, Class B: .868, 1.988, 2, p=.370, class C.871, 1.789, 2, .409)

for possible trends reflecting patterns found in the analyses of other measures (table 6.17). They suggested that between pre and dp tests the effect of group may be approaching significance.

In class C planned contrasts suggested that the sources of the difference found in the repeated measures ANOVA were between pre and post and between pre and dp tests, but not between post and dp tests. Again, this is perhaps surprising given the production practice class C experienced between post and dp test, yet not between the pre and post tests. The improvement between the pre and post tests may suggest a test effect (which may also have affected class A's scores, though not class B's as the EI group showed no improvement, see section 6.2.7).

It is noted that the results from these parametric tests may have been affected by the small and unequal sample sizes and the non-normal distribution of the data.

6.2.6 Non-parametric tests

The non-parametric tests generally supported the trends suggested by the analyses presented above. In the dataset as a whole, it was found that there were significant differences between pre and post and between pre and dp tests, but not between post and dp tests. This was also the case for the merged EI and PI groups and for classes A and C (perhaps surprising given the output practice experienced by learners in classes A and C between post and dp test). The intact class B's scores did not show any differences between any of the tests (the paired Wilcoxon tests suggested that there may have been a trend towards gains between the pre and post and between pre and dp tests). In class A, the EI group made pre-post and pre-dp test gains, and the PI group made pre-dp test gains and their pre-post gains were approaching statistical significance. There were no significant post-dp test differences in either the PI or EI groups in class A. In class B, the EI group made no gains between any of the tests whereas the PI group, in line with other measures, made gains between the pre and post tests and maintained these gains at dp test. There was no difference between the post and dp test scores.

6.2.7 Discussion and analysis of class C, the non-active control

As with other measures, a comparison of pre-post test gains in classes A and C was carried out to explore whether a school effect can explain the gains made by EI learners in class A but not in class B. It was found that there was no difference in the effect size of the pre-post test gains made by classes A and C (0.01 sdu). An independent samples t test supported this, as shown in table 6.18. It is noted that this is the only measure where there were no differences between the pre-post test gains made by classes A and C. One possible explanation for this is, as mentioned in section 4.1.2 a), is that during some of the narrative oral tests the learners from class A were taken from a lesson which involved explicit presentation and practice of the perfect tense. It is therefore possible that in class A, both PI and EI learners' pre test performances over-represented their developing systems, making gains at post test more difficult to demonstrate. However, as there was no systematic monitoring of evidence of this (e.g. pausing) during the tests, this explanation cannot be pursued further.

t	df ¹⁹	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
					Lower	Upper
.037	30	.971	.12875	3.52199	-7.06411	7.32161

Table 6.18 Independent samples t test to compare the gains made by classes A and C between pre and post speaking tests

Nonparametric tests were also carried out as the n was small and the distribution of data was not normal in some of the samples. The results are not presented here as there were no differences between the two analyses.

It is therefore acknowledged that gains made by PI and EI learners in class A between pre and post test were probably due to the test raising their awareness of the target features possibly due to the ethos of teach and test grammar in school 1. It is emphasised that no such effect was noted in school 2 as only one group of learners made any gains.

¹⁹ Equality of variances could be assumed

6.2.8 Summary of 6.2 Analysis of the speaking tests

The analyses in this section have indicated that in class B the PI learners made significant gains between the speaking pre and post tests, and maintained these gains at dp test, whereas the EI learners in class B made no gains. In class A, both groups of learners made significant gains between pre and post tests and maintained these gains at dp test. However, it cannot be claimed that these gains were due to the experimental intervention as class C also made similar gains without any FonF/S between the pre and post tests. Despite the FonF/S (fairly mechanical output practice) in classes A and C between the post and dp tests, no gains were made in this time by classes A and C or by the merged PI and EI groups.

6.3 Summary of chapter 6 and discussion

This chapter has presented a range of analyses which have suggested that for the writing and speaking measures PI had a positive impact compared to EI in class B but not in class A, in line with the listening and reading measures in chapter 5. In class B, learners receiving PI maintained their learning gains after 14-16 weeks. Equivalent EI learners in class B made no learning gains during the study and it therefore seems likely that they were unable to detect verb inflections whilst also interpreting items of higher CV in the EI activities. It was suggested the impact of PI and EI in class B was similar on both the sentence and narrative level oral production measures.

For the written production in class A both the PI and EI learners seemed to benefit from the experimental intervention as the non-active control did not make statistically significant pre-post test gains. However, the impact of PI and EI in class A on the speaking scores was difficult to ascertain as the non-active control made statistically similar pre-post test gains, possibly due to a test effect brought about in school 1 by the ethos of teach and test grammar in that school.

The pre-post test gains in the writing measure made by the EI learners in class A may have been because they were able to detect the target verb inflections incidentally to

the task that forced their attention on lexical items and/or overall sentential meaning in the input. In terms of IP theory, this may have been because they had sufficient attentional resources as their developing systems seemed already to have incorporated some representation of verb inflections (as measured by their interpretation and production of verb inflections at pre test). More attentional resources may also have been available to learners in class A than in class B because the former may have had better metalinguistic knowledge and/or vocabulary range (see sections 4.1 and 4.2). It is also acknowledged that the EGI at the start of each unit of work, combined with the input flood of target forms in the EI, may have been sufficient to reactivate their language knowledge / use in order to make greater gains between pre and post writing tests than the non-active control.

This explanation is probably not appropriate for the oral measure as differences between classes A and B's oral pretest scores were not evident (see section 4.2). However, it is still argued that the EI learners in class B may not have made gains because they did not detect the verb inflections in the EI materials due to their low CV as hypothesised by IP theory and as suggested in the other measures. In addition, for the oral production measure only, it is suggested that EI learners in class A made gains only because of the background ethos of teach and test grammar at school 1 (as demonstrated by class C's statistically similar gains) and not due to being able to attend in some way to verb inflections as a result of a higher developmental stage. This could be used to support the previous arguments regarding the role of developmental stages in input processing: oral production was the only measure not to reflect different developmental stages and it is also the only measure in which EI learners in both classes A and B did not make gains attributable to the EI²⁰.

As argued in section 5.3, general teaching in class A outside the experimental protocol may (also) have favourably affected learners' attention, enabling EI learners in class A but not in class B to make gains between the pre and post writing tests (and speaking tests, if these gains are not solely attributable to the school effect).

²⁰ This is not to suggest that the PI learners' gains in class A in oral production are attributable to PI: both PI and EI gains in oral production in class A could be attributed to the background ethos of teach and test grammar in school 1.

The style / sequencing / nature of the activities in class A may have made the EI learners in class A more likely to attend to (in a general sense or 'detect' in terms of IP) the input flood of verb inflections in the activities following an EGI. In contrast, EI learners in class B may not have been accustomed to planned sequences of activities focussing on one or two language forms and/or being expected to attend to features in input beyond the minimum required by the task.

Chapter 7 Summary and Conclusions

7.1 Summary of the study

The study presented in this thesis was a quasi-experiment carried out in two secondary schools in England amongst 86 year 9 learners of French. The study compared two input-based approaches: Processing Instruction (PI) and Enriched Input (EI) for teaching French verb inflections for tense (present and perfect), person and number.

This comparison largely took its theoretical foundation from Input Processing (IP) theory. PI was considered to be a reasonably faithful classroom operationalisation of this theory. According to IP theory, learners tend not to process apparently redundant language items (verb inflections, in this case) for meaning. The teaching materials in this study were designed to test this claim. The PI tasks were designed to force the learner to interpret the meaning of verb inflections. EI tasks allowed learners to interpret only the lexical items or overall sentential meaning i.e. learners did not have to interpret the verb inflections – if they did, it would have been incidental to the task set. In this sense, these activities resembled many current classroom reading and listening activities.

Attempts were made to strike a balance between maintaining the external and internal validity of the quasi-experiment. Two comparisons of PI and EI were carried out in two classes (A and B) from different schools; each class was divided into equivalent PI and EI groups using randomised matched pairs; a non-active control class (class C), parallel to the experimental class in school 1, also undertook the pre, post and delayed post (dp) tests; a wide range of verb inflections were used; a possible teacher bias was compensated for by alternately teaching EI and PI, and was also assessed in an interview; the 12 hours of experimental intervention were spread over 7 weeks; the possible influence of pupil attitudes to the experimental intervention was assessed; and the regular teaching and learning of French was

monitored to give some idea of what pupils were accustomed to, particularly in terms of grammar pedagogy.

Four measures of progression in the learning of the verb inflections were used: listening, reading, writing and speaking. A battery of tests was used to measure learning gains between the pre and post tests (conducted in the two weeks after the end of the intervention) and at dp tests (conducted 14 - 16 weeks after the end of the intervention). The written production measure included some narrative tasks and the oral production measure included some semi-spontaneous tasks (narratives and guided conversations).

The lessons in each class were monitored prior to and throughout the study, partly to assess whether any teaching of verb inflections may interfere with the internal validity of the study. Learning gains between pre and post tests cannot have been due to subversion of the experimental teaching protocol i.e. there was very little FonF/S outside the experimental treatment between pre and post tests, therefore differences between PI and EI learners' scores at post test were likely to be due to differential impacts of these teaching materials. The improvements made by class A and C learners between post and dp tests were likely to be mainly due to the FonF/S they experienced after the post tests.

7.2 Original features of the study

There were several features of this study which are original for PI research and which rendered this study more ecologically and internally valid than previous PI studies (and many Focus on Form studies): the range of target features, the duration of the intervention, the use of two different 'populations' of learners, the use of oral narratives, the length of time between the post and dp tests, the comparison with another input based approach, the use of the same explicit grammar instruction (EGI) with both experimental groups and the extensive monitoring of the teaching before, during and after the experimental intervention.

7.3 Findings

When the scores from the instructional groups in each class were merged (to form one merged PI group and one merged EI group), it was found that both groups made statistically significant gains and the type of input instruction alone did not have any statistically significant impact on the learners' scores in the four measures.

However, it was found that PI and EI tended to have differential impacts depending on the class learners were in.

In class A it was found that both PI and EI learners made similar learning gains in all the measures between pre and post tests and maintained (or increased) these at dp test.

In class B it was found that only the PI learners made statistically significant gains in all measures between pre and post tests and maintained these at dp test (though there was no FonF/S of the target forms between post and dp test). However, the EI learners made no progress in the learning of the target verb inflections during the study.

In class C learners made some statistically significant learning gains between pre and post tests in listening, reading and speaking measures. Their gains in the writing scores between pre and post test were not statistically significant.

In terms of the research questions, the findings can be summarised as follows:

- 1) Do structured input activities have any favourable effects compared to Enriched Input on year 9 learners' ability to *interpret* in reading and / or aurally French present and perfect tense verb inflections, as measured by a battery of achievement tests?

Yes, this was found for certain year 9 learners.

2) Do structured input activities have any favourable effects on year 9 learners' ability to *produce* in writing and / or orally French present and perfect tense verb inflections, as measured by a battery of achievement tests?

Yes, this was found for certain year 9 learners.

3) Are the same results for questions 1 & 2 maintained in *delayed* post tests, taken between 4 and 6 months after the start of the intervention?

Yes, this was found for certain learners.

4) Are the same results for questions 1, 2 & 3 obtained with learners from two 'similar' classes from different schools?

No, it was found that PI consistently had an advantage over EI in one class, but not in another.

7.4 Discussion of the findings

In class B, the improvement in all measures in the PI group was unlikely to be due to a test effect or to the EGI as the EI group made no gains. It was considered that the cause of the PI group's learning gains was most likely to be the structured input activities. The results from class B can therefore be used to support a role for PI in developing a verb inflection system in French for these learners. They also appear to support IP principle 1 in that these learners did not appear to process items of low communicative value (verb inflections) when they were embedded in input in which items of higher communicative value could be comprehended. This could offer evidence to suggest that attentional resources in second language learning are allocated according to some tension between items of higher / lower communicative value as defined by VanPatten i.e. that when attention is allocated to processing lexical items or overall sentential meaning, this is to the detriment of the processing of redundant verb inflections.

The following factors may explain why learning gains were made by EI learners in class A and not by those in class B.

i) *School / test effect*

It may be that a ‘school / test effect’ enabled learners in class A to make improvements. It was seen that class C, equivalent at pre test and in other relevant ways to class A, also made statistically significant gains between the pre and post tests in the listening, reading and speaking measures (despite having had no exposure to FonF/S in the target forms during this period). This school had an ethos of teaching and testing grammar, with a particular focus on verb inflections, and it is possible that the learners’ awareness / knowledge / competence in the target features was reactivated and rehearsed in the pre and post tests.

However, class C’s pre-post test gains in the writing measure were not statistically significant and so the improvement seen in learners in class A in this measure cannot be explained by a test/school effect. Furthermore, the test/school effect cannot account for the *extent* of the gains made by learners in class A in the reading and listening measures: the pre-post test gains made by class A were statistically significantly *greater* than class C’s. This suggests that the learners in class A benefited in some way from the PI and EI intervention.

However, it is acknowledged that the pre-post test gains made by PI and EI learners in class A in the oral production measure may have been entirely due to the school effect, as class C also made statistically similar gains.

ii) *Incidental input processing*

It is argued, then, that the experimental intervention (both PI and EI) was at least a partial cause of class A’s learning gains in the listening, reading and writing measures. This, in turn, suggests that EI learners in class A may have been attending to in some way (possibly interpreting the meaning of) verb inflections incidentally to carrying out the tasks which required comprehension of lexical items or sentential meaning (i.e. in terms of IP theory, processing form *and* meaning simultaneously).

IP theory provides some explanation for these apparently contradictory findings in sub principle 1e: "The Availability of Resources Principle: For learners to process either redundant meaningful grammatical forms or nonmeaningful forms, the processing of overall sentential meaning must not drain available processing resources" (VanPatten 1996 & 2004). Although classes A, B and C were all 'top ability' year 9 classes from similar schools, their pre test scores were actually found to be from different 'populations' for the listening, reading and writing measures i.e. classes A and C's (in school 1) scores were statistically significantly higher than class B's (in school 2). The learners in class A (and C) were at a higher developmental stage than learners in class B in terms of the interpretation and written production of verb inflections (and possibly also in terms of metalinguistic knowledge and/or vocabulary range). The EI learners in class A may therefore have had sufficient attentional resources to interpret communicatively redundant items in addition to comprehending lexical items or sentential meaning in the EI activities. This practice may have promoted learning gains in the interpretation and written production tests. In contrast, EI learners in class B may not have had sufficient processing resources to detect the verb inflections as well as completing the task set. This study could suggest that the higher developmental stage in the interpretation and written production of verb inflections (and possibly in their metalinguistic awareness and/or wider vocabulary range) of learners in class A meant that they were able to benefit from EI in ways that learners in class B were not. Additional support for this was that the oral production measure was the only measure in which classes A and B did *not* start at statistically significantly different stages and it was the only measure in which the EI learners in both classes A and B did not make any gains attributable to the EI activities.

This study cannot comment further on the existence /nature of this incidental processing: an investigation of the processes involved during the EI activities would involve a controlled laboratory experiment using measurements of implicit learning e.g. priming tasks.

Nevertheless, this study could suggest that when learners are at a higher developmental stage (though it cannot be said conclusively whether this is in terms

of their verb inflection system and /or lexicon and/or metalinguistic knowledge) a more incidental mode of using input e.g. tasks which engage the learner with more holistic sentence processing accompanied by a brief metalinguistic explanation, can be as effective as PI in terms of the development of the verb inflection system. In contrast, for learners who are at an earlier stage of developing a verb inflection system (and/or lexicon and/or metalinguistic knowledge) PI offers a useful approach to promoting learning of a verb inflection system.

iii) Role of the EGI

The claim that learners were engaging in incidental input processing is, however, not conclusive as it is possible that the brief EGI at the start of each unit may have been the cause of class A EI (and possibly PI) learners' gains between pre and post test, particularly given the ethos of explicit teaching and testing grammar in this school. This could explain the additional pre-post test gains made by class A compared to class C (which did not receive the EGI). However, the EGI was unlikely to have such an impact given that it was very brief, and several PI studies and many SLA researchers have dismissed the role of EGI alone in promoting language learning (see section 1.3.2 d). Nevertheless, it may be that the *combination* of EGI and EI served to promote learning, for example by raising learners' awareness of features in the input and increasing the likelihood that they attend in some way to them. This may have been more helpful for the learners in class A because they had already received considerable FonF/S regarding the relevant target features and/or because they were developmentally more ready as seen in their interpretation and production of verb inflections in the pre tests. It is acknowledged therefore that the use of the EGI with both groups, although strengthening the internal validity of the experiment in some respects (i.e. enabling the assignation of causality in class B to the structured input activities), has meant that the role of EGI in class A's groups remains unspecified. It is also recognised that this study cannot comment on whether metalinguistic knowledge was being activated during the test measurements or the input instruction tasks.

iv) Grammar pedagogy tactics prior to the study

A further reason for EI learners in class A making learning gains yet not those in class B is that general teaching characteristics in class A may have favourably affected whether / how learners' attended to (in a general sense, or interpreted the meaning of, in terms of IP) the target features embedded in the EI activities.

Although the conditions during the actual intervention in both classes were apparently similar (see chapter 4 – all materials were completed and learners in both classes *appeared* to be equally engaged in the tasks), the generally tighter approach to classroom management and / or sequencing of activities with a specific grammatical focus in class A prior to the study may have given learners general strategies and / or expectations regarding language learning which were not held by learners in class B. For example, learners in class A may have had greater propensity to make efforts to interpret more of the input than the minimum required by the task, given that they were accustomed to being asked about / tested on a small set of language features, verb inflections in particular, in many of their activities. Learners in class A may have expected, in a general way, that the EGI and subsequent activities would be related to each other and therefore make attempts to attend, in some way, to the features presented in the EGI in the tasks following it. The portrait of lessons in class B prior to and after the experimental intervention suggested that this pattern of teaching was not followed – although explicit grammar explanations were frequent, they were generally unpredictable in terms of their focus and the work preceding or following them, probably due to their reactive nature. EI learners in class B may therefore have been less likely to expect to have to attend, in some way, to verb inflections in activities following the EGI. These scenarios are offered not as conclusive explanations but to suggest that general teaching techniques prior to the study may have had an impact on the way learners processed the input in the experimental intervention and, therefore, on the test scores. It is acknowledged that such an explanation would require certain assumptions regarding the role of voluntary / conscious / effortful attention in second language learning, upon which this thesis cannot comment.

It is possible that a combination of the above (i – iv) explain why the EI learners in class A but not in class B made learning gains between the pre and post tests. This

demonstrates both the strengths of this study (in that a range of explanations has been found) and a weakness of field experiments (that absolute elimination of extraneous variables is not possible).

This study can therefore contribute to methodological debates regarding experimental educational research. It has highlighted the need for educational experiments to be carried out in a range of institutions, as important pedagogical and theoretical implications emerged from different schools. In addition, it has pointed to the need for close monitoring of the teaching prior to and during the experimental intervention, as this offered possible explanations for some of the findings in this study.

To address the ethical concerns of experimental educational research, some discussion is required regarding the finding that the EI group in class B did not make any gains in the interpretation or production of verb inflections throughout the study. First, it is suggested, as argued in section 3.1.2, that the random assignment of some learners in class B to PI and some to EI is comparable to the essentially random assignment, inherent in the school system, of pupils to different quantities and qualities of MFL grammar pedagogy. Secondly, as the EI learners did not make any gains at dp test once normal lessons had resumed, it is unlikely that the EI experimental intervention deprived them of otherwise beneficial teaching in the language features targeted in this study. Thirdly, it is argued that as the EI learners in class A did make learning gains, the EI materials themselves were only a less favourable alternative to PI given the context of class B. Finally, although it was established that neither instructional type had a clear impact on the number of verbs produced by learners (sections 6.1 and 6.2), it is acknowledged that more rigorous measurement of vocabulary development is required, as EI may have had beneficial effects in this area.

7.5 Limitations and further implications for future research

This study has made an important and unique contribution to PI research in that it has enabled an investigation of the theoretical claims upon which PI is based, and to

some extent supports them, by comparing PI with another input-based approach. However, this study has several limitations which require further research.

- i) The language sample included only verb inflections - further studies would be required to assess the impact of PI on other language features which may be less transparent and/or non-meaningful e.g. sentence syntax or gender concord.
- ii) The EI approach developed here was only one possible alternative to PI, and other studies are required to compare a range of input-based teaching approaches to strengthen the claim that PI can have advantages over a more incidental mode of processing input. In addition, rigorous measurement is required of any potential benefits of the comparison treatment (e.g. although some attempt was made in this study to compare the efficacy of both teaching approaches for the learning of lexical items, the measurements could have been improved).
- iii) This study suggested, in line with claims from other researchers, that the developmental stage of the learner may influence how input is processed. This points to the need for studies which make rigorous measurement of 'developmental stage' and explore input processing in learners from a range of stages.
- iv) The study suggested that some 'incidental' learning of verb inflections may have taken place. However, this was inconclusive as other factors could partially explain the results. Therefore, laboratory studies would be required to explore the occurrence / nature of incidental learning in activities such as EI amongst learners of different developmental stages using measures of implicit learning and memory, such as priming tasks.
- v) In terms of implications for more pedagogically-oriented research, although it was suggested that PI materials could be useful for the teaching and learning of verb inflections, more data should be collected and analysed to assess the affective impact of the materials and how they may fit into the MFLs curriculum. While some such data was collected for this study, it was beyond the scope of this thesis to analyse it in sufficient detail for these purposes.

vi) This study has demonstrated that the international FonF/S research agenda can be usefully explored in the context of MFLs teaching and learning in UK secondary schools. It is therefore suggested that further such studies, involving other FonF/S techniques could contribute to informing grammar pedagogy at this level.

vii) This study has, in turn, informed FonF/S research in various respects, with several implications for future grammar pedagogy research, particularly the importance of exploring FonF/S options at different levels and in different, carefully described contexts and the importance of taking measures to improve and systematise internal validity in FonF/S studies (e.g. by using semi-spontaneous oral production, administering delayed post tests after several months and controlling the teacher variable).

References

Adinolfi, N. 1994, *Prefabrication in Instructed Language Learning: A Study of Question Forms*, MA Dissertation, University of Southampton.

Aitchison, J. 1998, *The Articulate Mammal*, 4th edn, Routledge.

Alanen, R. 1995, "Input enhancement and rule presentation in second language acquisition.", in *Attention and Awareness in Foreign language Learning*, R. Schmidt, ed., University of Hawai'i Press, pp. 259-302.

Alderson, J. 2000, *Assessing Reading* Cambridge University Press.

Alderson, J., Clapham, C., & Steel, D. 1998, "Metalinguistic knolwedge, language aptitude and language proficiency", *Language Teaching Research*, vol. 1, no. 2, pp. 93-121.

Aljaafreh, A. & Lantolf, J. 1994, "Negative feedback as regulation and second language learning in the zone of proximal development", *The Modern Language Journal*, vol. 78, no. 4, pp. 465-483.

Allen, L. 2000, "Form-meaning connections and the French causative", *Studies in Second Language Acquisition*, vol. 22, pp. 69-84.

Allford, D. 2003, "'Grasping the nettle': aspects of grammar in the mother tongue and foreign languages", *Language Learning Journal*, vol. 27, pp. 24-32.

Andersen, R. 1983, "Transfer to Somewhere," in *Language Transfer in Language Learning*, S. Gass & L. Selinker, eds., Newbury House, pp. 177-201.

Andersen, R. & Shirai, Y. 1994, "Discourse motivations for some cognitive acquisition principles", *Studies in Second Language Acquisition*, vol. 16, pp. 133-156.

Anderson, J. 1995, *Learning and Memory: An Integrated Approach* John Wiley.

Atkinson, E. 2000, "Critical dissonance and cirical schizophrenia: the struggle between policy delivery and policy critique", *Research Intelligence*, vol. 73, pp. 14-17.

Bachman, L. & Palmer, A. 1996, *Language Testing in Practice* Oxford University Press, Oxford.

Baddeley, A., Gathercole, S., & Papagno 1998, "The phonological loop as a language learning device", *Psychological Review*, vol. 105, pp. 158-173.

Barcroft, J. & VanPatten, B. 1997, "Acoustic salience of grammatical forms: the effect of location, stress, and boundedness on Spanish L2 Input Processing," in *Contemporary Perspectives on the Acquisition of Spanish, volume 2: Production, Processing, and Comprehension*, W. Glass & A. Pérez-Leroux, eds., Cascadilla Press, pp. 109-121.

Bardovi-Harlig, K. 1992, "The use of adverbials and natural order in the development of temporal expression", *International Review of Applied Linguistics*, vol. 30, pp. 299-320.

Bardovi-Harlig, K. 2000, "Tense and aspect in second language acquisition: form, meaning, and use", *Language Learning*, vol. 50, no. Supplement 1.

Bates, E. & MacWhinney, B. 1989, "Functionalism and the competition model," in *The Cross-Linguistic Study of Sentence Processing*, B. MacWhinney & E. Bates, eds., Cambridge University Press, pp. 77-117.

Beard, R. & Willcocks, J. 2001 "The National Literacy Strategy in England: A major shake-up", Paper presented at the BERA Annual Conference, University of Leeds.

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*, vol. 5, no. 2, pp. 95-127.

Benati, A. 2003, "Explicit information vs. structured input activities on the acquisition of adjective agreement in Italian.", *under review*.

Benati, A. 2004, "The effects of structured input activities and explicit information on the acquisition of the Italian future tense," in *Processing Instruction: Theory, Research, and Commentary*, B. VanPatten, ed., Lawrence Erlbaum.

Berne, J. 1989 The effect of increased comprehensibility on learner's ability to consciously attend to content and form. Unpublished manuscript, University of Illinois at Urbana-Champaign.

Bernhardt, E. 1987, "Cognitive processes in L2: an examination of reading behaviours," in *Research in Second Language Learning: Focus on the Classroom*, J. Lantolf & Labarca, eds..

Bernhardt, E. 1991, *Reading Development in a Second Language* Ablex Publishing Corporation.

Blau, E. 1990, "The effect of syntax, speed and pauses on listening comprehension", *TESOL Quarterly*, vol. 24, pp. 746-753.

Blunkett, D. 2000 Influence or irrelevance? The Times Higher Education Supplement , 36-37. 2-4-2000.

Borg, S. 1996, "Language pedagogy and linguistics", *Language Awareness*, vol. 5, no. 2, pp. 118-124.

Borg, S. 1998, "Data-based teacher development", *ELT Journal*, vol. 52, no. 4, pp. 273-281.

Borg, S. 1999, "Teachers' theories in grammar teaching", *ELT Journal*, vol. 53, no. 3, pp. 157-167.

Boyd, S. 2001, *Teaching Modern Languages: Policy and practice in England, Wales and Northern Ireland* NFER.

Bransdorfer, R. 1989 "Processing function words in input: Does meaning make a difference?", Paper presented at the annual meeting of the American Association of Teachers of Spanish and Portuguese, San Antonio.

Brindely, G. 1998, "Assessing Listening Abilities", *Annual Review of Applied Linguistics*, vol. 18, pp. 171-191.

Broadbent, D. 1958, *Perception and Communication* Pergamon.

Brown, J. D. 1988, *Understanding research in second language learning: A teacher's guide to statistics and research design*. Cambridge University Press, London.

Brumfit, C., Mitchell, R., & Hooper, J. 1996, ""Grammar", "language" and classroom practice," in *Teaching and Learning in Changing Times*, M. Hughes, ed., Basil Blackwell, pp. 70-87.

Bryman, A. & Cramer, D. 1995, *Quantitative Data Analysis for Social Scientists* Routledge.

Bryman, A. & Cramer, D. 1997, *Quantitative Data Analysis for SPSS with Windows* Routledge.

Budge, D. 2002, "Launch of the First EPPI-Centre Reviews on 24 June 2002", *Research Intelligence*, vol. 80, pp. 28-29.

Burrell, G. & Morgan, G. 1979, *Sociological Paradigms and Organizational Analysis* Heinemann Educational Books.

Burstall, C. et al. 1974, *Primary French in the Balance* NFER, Windsor.

Bygate, M., Skehan, P., & Swain, M. 2001, *Researching Pedagogic Tasks: Second Language Learning, Teaching and Learning* Longman.

Bygate, M., Tonkyn, A., & Williams, E. 1994a, *Grammar and the Language Teacher* Prentice Hall.

Bygate, M., Tonkyn, A., & Williams, E. e. 1994b, *Grammar and the Language Teacher* Prentice Hall.

Cadierno, T., Glass, W., Lee, J., & VanPatten, B. 1991 *Processing tense in second language input: Lexical cues versus grammatical cues*. Unpublished manuscript The University of Illinois at Urbana-Champaign.

Campbell, D. & Stanley, J. 1963, *Experimental and Quasi-experimental Designs for Research* McNally.

Carr, T. & Curran, T. 1994, "Cognitive factors in learning about structured sequences", *Studies in Second Language Acquisition*, vol. 16, no. 2, pp. 205-230.

Carroll, S. 2001, *Input and Evidence: the raw material of second language acquisition* John Benjamins.

Chambers, F. & Richards, B. 1995, "The 'free conversation' and the assessment of oral proficiency", *Language Learning Journal*, vol. 11, pp. 6-10.

Chambers, G. 1999, *Motivating Language Learners* Multilingual Matters.

Charis 1996, *Français: Unit 1-5* The Association of Christian Teachers.

Chaudron, C. 1985, "Intake: On models and methods for discovering learners' processing of input", *Studies in Second Language Acquisition*, vol. 7, pp. 1-14.

Clark, J. L. 1988, *Curriculum Renewal in School Foreign Language Learning* Oxford University Press.

Clarke, E. 1985, "The acquisition of Romance, with special reference to French," in *The Crosslinguisitic Study of Language Acquisition vol 1: The Data*, D. Slobin, ed., Lawrence Erlbaum, pp. 687-782.

Cohen, J. 1977, *Statistical power analysis for the behavioural sciences*, Rev. ed. edn, Academic Press, New York.

Cohen, J., Dunbar, K., & McClelland, J. 1990, "On the control of automatic processes: A parallel distributed processing account of the Stroop effect", *Psychological Review*, vol. 97, pp. 332-361.

Cohen, L., Manion, L., & Morrison, K. 2000, *Research Methods in Education*, 5th edn, Routledge Falmer, London.

Collentine, J. 2004, "Commentary: Where processing-instruction research has been and where it should be going," in *Processing Instruction: Theory, Practice and Commentary*, B. VanPatten, ed., Lawrence Erlbaum.

Cook, G. 2001, "Classroom ritual and play: taking task to task", *Paper presented at AAAL, St.Louis 2001*.

Cooper, P. & McIntyre, D. 1996, *Effective Teaching and Learning* Open University Press.

Corder, S. 1978, "Language-learner language," in *Understanding Second nd Foriegn Language Learning: Issues and Approaches*, J. Richards, ed., Newbury House, pp. 71-93.

Coyle, D. 2000, "Meeting the challenge: developing the 3Cs curriculum," in *New Perspectives on Teaching and Learning Modern Languages*, S. Green, ed., Multilingual Matters, pp. 158-182.

Coyle, D. 2001 "Adolescent voices taking control: applying microgenetic analysis to explore strategic 'moves' made by learners in modern foreign language and bilingual learning contexts", Paper presented at the BERA Annual Conference at the University of Leeds.

Creemers, B. 1994, "Effective instruction: an empirical basis for a theory of educational effectiveness," in *Advances in School Effectiveness Research*, D. Reynolds & B. Creemers, eds., Pergamon.

Day, E. & Shapson, S. 1991, "Integrating formal and functional approaches to language teaching in French immersion: An experimental study.", *Language Learning*, vol. 41, pp. 25-58.

DeKeyser, R. 1998, "Beyond focus on form: cognitive perspectives on learning and practicing a second language grammar," in *Focus on Form in Classroom Second Language Acquisition*, C. Doughty & J. Williams, eds., Cambridge University Press, pp. 42-63.

DeKeyser, R., Salaberry, M., Robinson, P., & Harrington, M. 2002, "What gets processed in processing instruction: a commentary on Bill VanPatten's "Processing Instruction: Update""", *Language Learning*, vol. 52, no. 4, pp. 805-823.

DeKeyser, R. & Sokalski, K. 1996, "The differential role of comprehension and production practice", *Language Learning*, vol. 46, pp. 613-642.

DfEE 1999, *Modern Foreign Languages: The National Curriculum* DfEE.

DfEE 2001, *Modern Foreign Languages: A response by the Department for Education and Employment to the final report and recommendations of the Nuffield Languages Inquiry*, <http://www.dfee.gov.uk/nuffield/> DfEE.

DfES 2003, *Framework for teaching modern foreign languages: Years 7, 8 and 9* DfES.

Diamond, I. & Jeffries, J. 2001, *Beginning Statistics* Sage.

Dienes, Z., Broadbent, D., & Berry, D. 1991, "Implicit and explicit knowledge bases in artificial grammar learning", *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol. 17, pp. 875-887.

Dickson, P. 1996, *Using Target Language: a View from the Classroom* National Foundation for Educational Research.

Dietrich, R., Klein, W., & Noyau, C. 1995, *The Acquisition of Temporality in a Second Language* John Benjamins.

Dobson, A. 1998, *MFL Inspected: Reflections on inspection findings 1996-7* CILT.

Dörnyei, Z. 2001 Researching L2 Motivation: Where Are We Going and How Are We Doing It? Paper presented at LIE Research Seminar, University of Southampton

Doughty, C. 1991, "Second language instruction makes a difference: Evidence from an empirical study on relativization", *Studies in Second Language Acquisition*, vol. 13, pp. 431-469.

Doughty, C. 2001, "Cognitive underpinnings of focus on form," in *Cognition and Second Language Instruction*, P. Robinson, ed., Cambridge University Press, pp. 206-257.

Doughty, C. & Varela, E. 1998, "Communicative focus on form," in *Focus on Form in Classroom Second Language Acquisition*, C. Doughty & J. Williams, eds., Cambridge University Press, pp. 114-139.

Doughty, C. & Williams, J. 1998, "Pedagogical choices in focus on form," in *Focus on Form in Classroom Second Language Acquisition*, C. Doughty & J. Williams, eds., Cambridge University Press, pp. 197-262.

Eimer, M., Nattkemper, E., Schröger, E., & Prinz, W. 1996, "Involuntary attention," in *Handbook of Perception and Action Volume 3 : Attention*, O. Neumann & A. Sanders, eds., Academic Press, pp. 155-184.

Elliott, J. 1991, *Action research for Educational Change* Open University Press, Buckingham.

Elliott, J. 2001, "Making Evidence-based Practice Educational", *British Educational Research Journal*, vol. 27, no. 5, pp. 555-574.

Ellis, N. 1993, "Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge", *European Journal of Cognitive Psychology*, vol. 5, pp. 289-318.

Ellis, N. 1999, "Cognitive Approaches to SLA", *Annual Review of Applied Linguistics*, vol. 19, pp. 22-42.

Ellis, R. 1990, *Instructed second language acquisition* Basil Blackwell.

Ellis, R. 1993, "The structural syllabus and second language acquisition", *TESOL Quarterly*, vol. 27, no. 1, pp. 91-113.

Ellis, R. 1994, *The study of second language acquisition*, 2nd edn, Oxford University Press.

Ellis, R. 1995, "Appraising second language acquisition theory in relation to language pedagogy," in *Principle and Practice in Applied Linguistics*, V. Cook & B. Seidlhofer, eds., pp. 73-90.

Ellis, R. 1999, "Input-based Approaches to Teaching Grammar: A Review of Classroom-Oriented Research", *Annual Review of Applied Linguistics*, vol. 19, pp. 64-80.

Farley, A. 2001, "Authentic processing instruction and the Spanish subjunctive", *Hispania*, vol. 84, pp. 289-299.

Field, A. 2000, *Discovering statistics using SPSS for Windows* Sage Publications, London.

Fitz-Gibbon, C. 1984, *Meta-Analysis: An Explication*, Curriculum, Evaluation and Management Centre, Durham, 47.

Fitz-Gibbon, C. 1996, *Monitoring Education Indicators, Quality and Effectiveness* Cassell.

Fitz-Gibbon, C. 1999, "Education: High potential not yet realized", *Public Money & Management*, vol. 19, no. 1, pp. 33-40.

Fitz-Gibbon, C. 2000, "Education: realising the potential," in *What Works?*, H.Davies, S.Nutley, & P.Smith, eds., The Policy Press, pp. 69-92.

Fitz-Gibbon, C. 2001, "What's all this about 'evidence'?", *Learning and Skills Research*, vol. 5, no. 1, pp. 27-29.

Fitz-Gibbon, C. & Reay, D. 1982, *Peer tutoring: brightening up FL teaching in an urban comprehensive school*.

Fodor, J. A. 1983, *The Modularity of the Mind: An essay on Faculty Psychology* MIT Press.

Fotos, S. 1993, "Consciousness-raising and noticing through focus on form: Grammar task performance vs. formal instruction", *Applied Linguistics*, vol. 14, no. 4, pp. 385-407.

Fotos, S. 1994, "Integrating grammar instruction and communicative language use through grammar consciousness-raising tasks", *TESOL Quarterly*, vol. 28, no. 2, pp. 323-351.

Freedman, E. 1978, *Experimentation into foreign language teaching methodology*.

Gall, M., Borg, W., & Gall, J. 1996, *Educational Research: An Introduction* Longman.

Glass, G. 1979, "Policy for the unpredictable", *Educational Researcher*, vol. 8, no. 9, pp. 12-14.

Glass, G., McGaw, B., & Smith, M. 1981, *Meta-analysis in social research* Sage Publications.

Gass, S. 1988, "Integrating research areas: a framework for second language studies", *Applied Linguistics*, vol. 9, pp. 198-217.

Gass, S. 1997, *Input, Interaction and the Second Language Learner* Lawrence Erlbaum.

Gathercole, S. & Baddeley, A. 1993, *Working Memory and Language* Erlbaum.

Glass, W. 1994, *What learners say they notice in the input* Paper delivered at the annual meeting of the American Association for Applied Linguistics, Baltimore.

Goodman, K. 1968, *The Psycholinguistic Nature of the Reading Process* Wayne State University Press.

Goodwin, W. & Goodwin, L. 1996, *Understanding Quantitative and Qualitative Research in Early Childhood Education* Teachers College Press.

Graham, S. 1997, *Effective Language Learning* Multilingual Matters.

Gray, C. 1999, "In defence of the secondary teacher? A PGCE tutor's reaction to the great grammar debate", *Language Learning Journal*, vol. 19, no. 40, p. 45.

Green, P. & Hecht, K. 1992, "Implicit and explicit grammar: an empirical study.", *Applied Linguistics*, vol. 13, pp. 168-184.

Grenfell, M. 2000, "Modern languages - beyond Nuffield and into the 21st Century", *Language Learning Journal*, vol. Winter, no. 22, pp. 23-29.

Grenfell, M. & Harris, V. 1999, *Modern Languages and Learning Strategies: In theory and practice* Routledge.

Griffiths, M. 1998, "Truths and Methods," in *Educational Research for Social Justice*, M. Griffiths, ed., Open University Press, pp. 33-43.

Hammersley, M. 1987, "Some notes on the terms 'validity' and 'reliability'", *British Educational Research Journal*, vol. 13, no. 1, pp. 73-81.

Hammersley, M. 1999, "Some Reflections on the Current State of Qualitative Research", *Research Intelligence*, vol. 70, pp. 16-18.

Hargreaves, D. 1997, "In defence of research for evidence-based teaching", *British Educational Research Journal*, vol. 23, no. 4, pp. 405-419.

Harley, B. 1986, *Age in Second Language Acquisition* Multilingual Matters.

Harley, B. 1989, "Functional grammar in French immersion: a classroom experiment", *Applied Linguistics*, vol. 10, no. 3, pp. 331-359.

Harley, B. 1993, "Instructional strategies and SLA in early French immersion", *Studies in Second Language Acquisition*, vol. 15, no. 2, pp. 245-259.

Harley, B. 1994, "Appealing to consciousness in the L2 classroom", *AILA Review*, vol. 11, pp. 57-68.

Harley, B. & Swain, M. 1978, "An analysis of the verb system used by young learners of French", *Interlanguage Studies Bulletin*, vol. 3, no. 1, pp. 35-79.

Harris, A. 1998, "Effective teaching: a review of the literature", *School Leadership and Management*, vol. 18, no. 2, pp. 169-183.

Harris, V., Burch, J., Jones, B., & Darcy, J. 2001, *Something to Say?* CiLT.

Hatch, E. 1983, "Simplified input and second language acquisition," in *Pidginization and Creolization as Language Acquisition*, R.W.Andersen, ed., Newbury House, pp. 64-88.

Hatch, E. & Lazaraton, A. 1991, *The research manual: Design and statistics for applied linguistics*. Newbury House, Rowley, MA.

Hatch, E., Shirai, Y., & Fantuzzi, C. 1990, "The need for an integrated theory: connecting modules", *TESOL Quarterly*, vol. 24, pp. 697-716.

Hawkins, E. 1996, *30 Years of Language Teaching CiLT*.

Hawkins, R. 1989, "Do second language learners acquire restrictive relative clauses on the basis of relational or configurational information? The acquisition of French subject, direct object, and restrictive relative clauses by second language learners", *Second Language Research*, vol. 5, no. 2, pp. 155-188.

Hawkins, R. 2002, "The role of morphology and syntax in second language development", *Paper presented at BAAL / CUP seminar, July 2002, University of Southampton*.

Henning, G. 1983, "Oral proficiency testing: comparative validities of interview, imitation, and completion methods", *Language Learning*, vol. 33, pp. 315-332.

Herschensohn, J. 2003, "Verbs and rules: two profiles of French morphology acquisition", *French Language Studies*, vol. 13, pp. 23-45.

Hillage, J., Pearson, R., Anderson, A., & Tamkin, P. 1998, *Excellence in Research on Schools*, Research Report RR 74, DfEE.

Hitchcock, G. & Hughes, D. 1995, *Research and the Teacher*, 2nd edn, Routledge.

Holmes, D. 2000, *Multivariate Data Analysis*, 3rd year undergraduate Social Statistics module, University of Southampton.

Housen, A. 1998, "It's about time: the development of tense and aspect in second language acquisition", *Paper presented at the Workshop on Bilingualism and Second Language Acquisition at the University of Southampton*.

Houston, T. 1997, "Sentence Processing in Spanish as a Second Language: a study of word order and background knowledge," in *Contemporary Perspectives on the Acquisition of Spanish. Volume 2: Production, Processing, and Comprehension*, W. Glass & A. Pérez-Leroux, eds., Cascadilla Press, pp. 123-134.

Howard, M. 1999, "The advanced learner: poor relation of the second language acquisition family?", *Association for French Language Studies Cahiers*, vol. 5, no. 1, pp. 7-26.

Howard, M. 2002, "Temporal reference in advanced French interlanguage", *Paper presented at BAAL / CUP seminar, July 2002, University of Southampton*.

Howell, D. 1997, *Statistical Methods for Psychology*, 4th edn, Duxbury, Belmont, CA.

Hudson, R. 2001, "Grammar teaching: why, when, how and what?", *LLAS Subject Centre Grammar Supplement* pp. 3-4.

Hulstijn, J. & DeKeyser, R. 1997, "Testing SLA theory in the research laboratory", *Studies in Second Language Acquisition*, vol. 19, no. 2.

Hunter, J. & Schmidt, F. 1990, *Methods of meta-analysis* Sage, Newbury Park, CA.

Hunter, J., Schmidt, F., & Jackson, G. 1982, *Meta-analysis: Cumulating Research Findings across Studies* Sage Publications.

Hurman, J. 1992, "Performance in the A level speaking test by candidates with GCSE training: oral examiners' views", *Language Learning Journal*, vol. 5, pp. 8-10.

Isaac, S. & Michael, W. 1995, *Handbook in research and evaluation : a collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences*, 3rd edn, EdITS, San Diego.

Izumi, S. 2002, "Output, input enhancement, and the Noticing Hypothesis: An experimental study on ESL relativization", *Studies in Second Language Acquisition*, vol. 24, no. 4, pp. 541-577.

Jacoby, L. 1983, "Remembering the data: Analyzing interactive processes in reading", *Journal of Verbal Learning and Verbal Behaviour*, vol. 22, pp. 485-508.

Johnson, K. 1994, "Teaching declarative and procedural knowledge," in *Grammar and the Language Teacher*, M. Bygate, A. Tonkyn, & E. Williams, eds..

Jones, F. 1992, "A language-teaching machine: Input, uptake and output in the communicative classroom", *System*, vol. 20, no. 2, pp. 133-150.

Jones, K. & Bryant, I. 1998, Quantitative Analysis in Education Research. RGSE, University of Southampton.

Jourdenais, R. 1998, "The effects of textual enhancement on the acquisition of the Spanish preterit and imperfect.", *Unpublished doctoral dissertation. Georgetown University, Washington, DC.*

Jourdenais, R., Ota, M., Stauffer, S., Boyson, B., & Doughty, C. 1995, "Does textual enhancement promote noticing? A think-aloud protocol analysis," in *Attention and awareness in foreign language learning*, R. Schmidt, ed., University of Hawai'i Press, pp. 183-216.

Jung, J. 1971, *The experimenter's dilemma* Harper and Row.

Just, M. & Carpenter, P. 1993, "A capacity theory of comprehension: Individual differences in working memory", *Psychological Review*, vol. 99, pp. 122-149.

Kanji, G. 1993, *100 Statistical Tests* Sage.

Kellerman, E. 1983, "Now you see it, now you don't," in *Language Transfer in Language Learning*, S. Gass & L. Selinker, eds., Newbury House, pp. 112-134.

Kember, D. 2000, *Action Learning and Action Research: Improving the Quality of Teaching and Learning* Kogan Page.

Kerlinger, F. 1970, *Foundations of Behavioral Research* Holt, Rinehart & Winston.

Klapper, J. 1997, "Language learning at school and university: the great grammar debate continues (1)", *Language Learning Journal*, vol. 16, pp. 22-27.

Klapper, J. 2003, "Taking communication to task? A critical review of recent trends in language teaching", *Language Learning Journal*, vol. 27, pp. 33-42.

Klein, W. 1986, *Second Language Acquisition* Cambridge University Press.

Krampsch, C. 1995, "The applied linguist and the foreign language teacher: can they talk to each other?", in *Principle and Practice in Applied Linguistics: Studies in Honour of H.G. Widdowson*, V. Cook & B. Seidlhofer, eds., Oxford University Press, pp. 43-56.

Krashen, S. 1982, *Principles and Practice in Second Language Acquisition* Pergamon.

Krashen, S. 1985, *The Input Hypothesis* Longman, London.

Krashen, S. & Terrell, T. 1983, *The natural approach: Language acquisition in the classroom* Alemany Press.

L'Huillier, M. 1999, *Advanced French Grammar* Cambridge University Press.

Labeau, E. 2002, "Le rôle de l'aspect lexical dans l'acquisition des temps du passé français par des apprenants anglophones avancés", *Paper presented at BAAL / CUP seminar, July 2002, University of Southampton*.

Lachman, R., Lachman, J., & Butterfield, E. 1979, *Cognitive psychology and information processing: An Introduction* Erlbaum.

Lagemann, E. 2000, *An Elusive Science: The Troubling History of Education Research* University of Chicago Press.

Larsen-Freeman, D. & Long, M. 1991, *An Introduction to Second Language Acquisition Theory and Research* Longman.

Lauer, P., Streby, W., & Battig, W. 1976, "The effects of alphabetic organization and delayed retention of semantically similar words", *Journal of Experimental Psychology: Human Learning and Memory*, vol. 2, pp. 182-189.

LeCompte, M. & Schensul, J. 1999, *Designing and Conducting Ethnographic Research* Altamira.

Lee, J. 1987, "The Spanish subjunctive: an information processing perspective", *The Modern Language Journal*, vol. 71, pp. 50-57.

Lee, J. 1998, *The invisible child* CILT, London.

Lee, J. 1999, "Five types of input and the various relationships between form and meaning," in *Form and Meaning: Multiple Perspectives*, J. Lee & A. Valdman, eds., Heinle & Heinle, pp. 25-42.

Lee, T. & Magill, R. 1983, "The locus of contextual interference in motor-skill acquisition", *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol. 9, pp. 730-746.

Leeman, J., Arteagoitia, I., Fridman, B., & Doughty, C. 1995, "Integrating attention to form with meaning: Focus on form in content-based Spanish instruction," in *Attention and awareness in foreign language learning*, R. Schmidt, ed., University of Hawai'i Press, pp. 217-258.

Leow, R. 1993, "To simplify or not to simplify: A look at intake", *Studies in Second Language Acquisition*, vol. 15, pp. 333-355.

Leow, R. 1997, "The effects of input enhancement and text length on adult L2 readers' comprehension and intake in second language acquisition", *Applied Language Learning*, vol. 8, pp. 151-182.

Lightbown, P. 1991, "What have we here? Some observations on the influence of instruction on L2 learning," in *Foreign/Second Language Pedagogy Research: A Commemorative Volume for Claus Faerch*, R. Phillipson et al., eds., Multilingual Matters.

Lightbown, P. 1992, "Getting quality input in the second / foreign language classroom," in *Text and Context: Cross-disciplinary and Cross-cultural Perspectives on Language Study*, C. Kramsch & S. McConnell-Ginet, eds., D.C. Heath, pp. 187-197.

Lightbown, P. 1998, "The importance of timing in focus on form," in *Focus on Form in Classroom Second Language Acquisition*, C. Doughty & J. Williams, eds., Cambridge University Press, pp. 177-196.

Lightbown, P. 2001 "Input filters in second language acquisition", in *Eurosla Yearbook*, Eurosla, Paderborn, Germany, pp. 79-97.

Lightbown, P., Halter, R., & White, J. 2002, "Comprehension-based learning: The limits of "do it yourself""", *Canadian Modern Language Review*.

Lightbown, P. & Spada, N. 1993, *How languages are learned* Oxford University Press.

Little, D. & Singleton, D. 1991, "Authentic texts, pedagogical grammar and Language Awareness in foreign language learning," in *Language Awareness in the Classroom*, C. James & P. Garrett, eds., Longman, pp. 123-132.

Long, M. 1985, "Input and second language acquisition theory," in *Input in Second Language Acquisition*, S.M.Gass & C.Madden, ed., Newbury House, pp. 377-393.

Long, M. 1991, "Focus on form: a design feature in language teaching methodology.", in *Foreign Language Research in Cross-Cultural Perspective*, K. de Bot, R. Ginsberg, & C. Kramsch, eds., Benjamins.

Long, M. & Robinson, P. 1998, "Focus on form: Theory, research, and practice," in *Focus on Form in Classroom Second Language Acquisition*, C. Doughty & J. Williams, eds., Cambridge University Press, pp. 15-41.

Lyster, R. & Ranta, E. 1997, "Corrective feedback and learner uptake: negotiation of form in communicative classrooms", *Studies in Second Language Acquisition*, vol. 19, pp. 37-61.

Macaro, E. 2001a, "Using the target language: reviewing the issues", *Links* no. 23, pp. 6-7.

Macaro, E. 2001b "Writing strategies in Modern Foreign Languages at Key Stage 4: description and preliminary findings of an interventionist study", Paper presented at the BERA Annual Conference at the University of Leeds.

Mackey, A. & Philp, J. 1998, "Conversational interaction and second language development: recasts, responses and red herrings?", *The Modern Language Journal*, vol. 82, no. 3, pp. 338-356.

Macrory, G. & Stone, V. 1996, "Knowing and using the perfect tense in French," in *Language and Education: Papers from the Annual Meeting of the British Association for Applied Linguistics* , G. Blue & R. Mitchell, eds., Multilingual Matters, pp. 107-118.

Macrory, G. & Stone, V. 2000, "Pupil progress in the acquisition of the perfect tense in French: the relationship between knowledge and use", *Language Teaching Research*, vol. 4, no. 1, pp. 44-82.

Mangubhai, F. 1991, "The processing behaviours of adult second language learners and their relationship to second language proficiency.", *Applied Linguistics*, vol. 12, pp. 268-297.

Marsden, E. 2001, "Implications of the developments in the National Curriculum for MFL and GCSE specifications", *Liaison: Subject Centre Grammar Supplement*, vol. 3, pp. 19-21.

McLaughlin, B. & Heredia, R. 1996, "Information-processing approaches to research on second language acquisition and use," in *Handbook of Second Language Acquisition*, W. Ritchie & T. Bhatia, eds., Academic Press, pp. 213-228.

McNab, R. 1994, *Avantage* Heinemann Educational.

McNab, R. & Barrabé, F. 1993, *Avantage 2* Heinemann, Oxford.

McNiff, J. 1993, *Teaching as learning: an action research approach* Routledge, London.

McPake, J. 1999, *Foreign Language in the Upper Secondary School: A study of the causes of decline*, Scottish Council for Educational Research, 91.

Metcalfe, P. 1997, *It's just a word: CALL, French verbs and mixed-ability pupils*, PhD thesis, The Open University.

Metcalfe, P., Laurillard, D., & Mason, R. 1995, "Decline of written accuracy in pupils' use of French verbs", *Language Learning Journal*, vol. 12, pp. 47-50.

Mitchell, R. 1988, *Communicative Language Teaching in Practice* CILT.

Mitchell, R. 1994, "Grammar, syllabuses and teachers," in *Grammar and the Language Teacher*, M. Bygate, A. Tonkyn, & E. Williams, eds., Prentice Hall, pp. 90-104.

Mitchell, R. 1996, "Language, education and applied linguistics in a changing world," in *Language, Education and Society in a Changing World*, Hickey & E. Williams, eds., pp. 7-20.

Mitchell, R. 2000a, "Applied linguistics and evidence-based classroom practice: the case of foreign language grammar pedagogy", *Annual Review of Applied Linguistics* pp. 281-303.

Mitchell, R. 2000b, "Tracing the link between teaching and learning in naturalistic research: the case of Modern Foreign Languages", *Paper presented at Research and Graduate School of Education research seminar, University of Southampton*.

Mitchell, R. 2003, "Rethinking the concept of progression in the National Curriculum for Modern Foreign Languages: a research perspective", *Language Learning Journal*, vol. 27, pp. 15-23.

Mitchell, R. & Brumfit, C. 2001, "The place of knowledge about language," in *Reflections on Language and Language Learning. In honour of Arthur van Essen*, M. Bax & J.-W. Zwart, eds., pp. 279-292.

Mitchell, R. & Dickson, P. 1997, "Progression in Foreign Language Learning", *Centre for Language in Education Occasional Paper No.45 University of Southampton*.

Mitchell, R. & Hogg, I. 2001 "Evaluation of 'Best Practice' Research in Modern Foreign Languages", Paper presented at the BERA Annual Meeting 2001, Symposium: Towards a Research Strategy for Teaching and Learning in Modern Foreign Languages.

Mitchell, R. & Hooper, J. 1992, "Teachers' views of language knowledge," in *Language Awareness in the Classroom*, C. James & P. Garrett, eds., Longman, pp. 40-50.

Mitchell, R. & Martin, C. 1997, "Rote learning, creativity and 'understanding' in classroom foreign language teaching", *Language Teaching Research*, vol. 1, no. 1, pp. 1-27.

Mitchell, R. & Myles, F. 1998, *Second Language Learning Theories* Oxford University Press.

Miyake, A. & Shah, P. 1999, *Models of Working Memory: Mechanisms of active maintenance and executive control* Cambridge University Press.

Moore, D. & McCabe, G. 1989, *Introduction to the practice of statistics* W. H. Freeman Co..

Musumeci, D. 1989, *The ability of second language learners to assign tense at the sentence level: A cross-linguistic study*, Doctoral, University of Illinois at Urbana-Champaign.

Myles, F. 1995, "Interaction between linguistic theory and language processing in SLA", *Second Language Research*, vol. 11, no. 3, pp. 235-266.

Myles, F. 2003, "The emergence of morpho-syntactic structure in French L2," in *Focus on French as a Foreign Language*, J. Dewaele, ed., Multilingual Matters.

Myles, F. (in press), "From data to theory: the over-representation of linguistic knowledge in SLA," in *Empirical evidence and theories of representation in current research in Second Language Acquisition*, R. Towell & R. Hawkins, eds., Transactions of the Philological Society.

Myles, F., Hooper, J., & Mitchell, R. 1998, "Rote or rule? Exploring the role of formulaic language in classroom foreign language learning.", *Language Learning*, vol. 48, no. 3, pp. 323-363.

Myles, F., Mitchell, R., & Hooper, J. 1999, "Interrogative chunks in French L2: A basis for creative construction?", *Studies in Second Language Acquisition*, vol. 21, pp. 49-80.

Naiman, N. 1974, "The use of elicited imitation in second language acquisition research", *Working Papers on Bilingualism*, vol. 2, pp. 1-37.

Naiman, N., Frohlich, M., Stern, H., & Todesco, A. 1978, *The Good Language Learner* OISE.

Neil, P. & Laverty, R. 2001, "An analysis of pupils' writing skills in the French GCSE examination", *Francophonie*, vol. 24, pp. 12-17.

Neumann, O. & Sanders, A. 1996, *Handbook of Perception and Action Volume 3: Attention* Academic Press Ltd.

Norris, J. & Ortega, L. 2000, "Effectiveness of L2 instruction: a research synthesis and quantitative meta-analysis", *Language Learning*, vol. 50, no. 3, pp. 417-528.

Oakley, A. 1998, "Experimentation and social interventions: a forgotten but important history", *British Medical Journal*, vol. 31, pp. 1239-1242.

OfSTED 1995, *Modern Foreign Languages: A Review of Inspection Findings 1993-1994*, HMSO.

OfSTED 2001, *The Annual Report of Her Majesty's Chief Inspector of Schools 1999-2000*, DfEE Stationery Office.

OfSTED 2002, *Ofsted Subject Reports: Secondary Modern Foreign Languages 2000-2001*, OFSTED.

Overstreet, M. 1998, "Text enhancement and content familiarity: The focus of learner attention", *Spanish Applied Linguistics*, vol. 2, no. 229, p. 258.

Pachler, N. 2003, "Foreign language teaching as an evidence-based profession?", *Language Learning Journal*, vol. 27, pp. 4-14.

Page, B. 1992, *Letting go - taking hold: a guide to independent language learning by teachers for teachers* CILT, London.

Page, E. 1999, *An investigation into the interlanguage of learners of French as a foreign language in a classroom context*, University of Southampton.

Papagno, C., Valentine, T., & Baddeley, A. 1991, "Phonological short-term memory and foreign-language vocabulary learning", *Journal of Memory & Language*, vol. 30, pp. 331-347.

Paulston, C. 1972, "Structural pattern drills: a classification," in *Teaching English as a Second Language*, H. Allen & R. Campbell, eds., MacGraw-Hill, pp. 129-138.

Perdue, C. 1993, *Adult language acquisition: cross-linguistic perspectives. Volume 2: The Results* Cambridge University Press.

Perkins, K. 1998, "Assessing Reading", *Annual Review of Applied Linguistics*, vol. 18, no. 208, p. 218.

Peters, A. 1985, "Language segmentation: operating principles for the perception and analysis of language," in *The cross-linguistic study of language acquisition, vol. 2: Theoretical Issues*, D. Slobin, ed., Erlbaum, pp. 1029-1067.

Pica, T. 1985, "Linguistic simplicity and learnability: Implications for syllabus design," in *Modelling and Assessing Second Language Acquisition*, K. Hyltenstam & M. Pienemann, eds., Multilingual Matters, pp. 137-152.

Pienemann, M. 1998, *Language Processing and Second Language Development: Processability Theory* John Benjamins, Amsterdam.

Pinker, S. 1982, "A theory of the acquisition of lexical interpretive grammars," in *The Mental Representation of Grammatical Relations*, J. Bresnan, ed., MIT Press, pp. 655-726.

Polio, C. & Gass, S. 1997, "Replication and reporting: a commentary.", *Studies in Second Language Acquisition*, vol. 19, pp. 499-508.

Popper, K. 1968, *The Logic of Scientific Discovery*, 2nd edn, Hutchinson, London.

Posner, M. & Snyder, C. 1975, "Attention and Cognitive Control," in *Information Processing and Cognition*, R. Solso, ed., Erlbaum, pp. 55-85.

Prévost, P. & White, L. 2000, "Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement.", *Second Language Research*, vol. 16, no. 2, pp. 103-133.

Pring, R. 2000, *Philosophy of educational research* Continuum, London.

QCA 2000a, *GCSE Criteria for Modern Foreign Languages*, HMSO.

QCA. 2000b *Schemes of Work for Key Stage 3 French*. QCA.

Radford, A. 1990, *Syntactic theory and the Acquisition of English Syntax: The Nature of Early Child Grammars of English* Blackwell.

Ranta, L. 1998, *Focus on Form from the Inside: The Significance of Grammatical Sensitivity for L2 Learning in Communicative ESL Classrooms.*, Unpublished doctoral dissertation, Concordia University Montreal.

Reber, A. 1967, "Implicit learning of artificial grammars", *Journal of Verbal Learning and Verbal Behaviour*, vol. 77, pp. 317-327.

Reber, A., Kassin, S., Lewis, S., & Cantor, G. 1980, "On the relationship between implicit and explicit modes in the learning of a complex rule structure", *Journal of Experimental Psychology: Human Learning and Memory*, vol. 6, pp. 492-502.

Rendall, H. 1998, *Stimulating grammatical awareness: A fresh look at language acquisition*.

Reynolds, D. & Mujis, D. 1999, "The effective teaching of mathematics: a review of research", *School Leadership and Management*, vol. 19, no. 3, pp. 273-288.

Robinson, P. 1995, "Attention, Memory, and the "Noticing Hypothesis""", *Language Learning*, vol. 45, no. 2, pp. 283-331.

Robinson, P. 1997, "Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions", *Studies in Second Language Acquisition*, vol. 19, pp. 223-247.

Rosa, E. & O'Neill, M. 1998, "Effects of stress and location on acoustic salience at the initial stages of Spanish L2 input processing", *Spanish Applied Linguistics*, vol. 2, pp. 24-52.

Rosenthal, R. 1991, *Meta-analysis Procedures for Social Research* Sage Publications, Beverly Hills.

Rudduck, J., McIntyre, D., & (eds.) 1998, *Challenges for Educational Research* Paul Chapman BERA.

Rule, S. & Marsden, E. 2004, "The status of functional categories in the acquisition of negation in early French L2 grammars.", *under review*.

Rutherford, W. & Sharwood Smith, M. 1985, "Consciousness raising and universal grammar", *Applied Linguistics*, vol. 6, pp. 274-282.

Salaberry, M. 1997, "The role of input and output practice in second language acquisition", *The Canadian Modern Language Review*, vol. 53, pp. 422-451.

Salaberry, M. 1998, "On Input Processing, True Language Competence, and Pedagogical Bandwagons: A Reply to Sanz and VanPatten", *The Canadian Modern Language Review*, vol. 54, no. 2, pp. 274-285.

Sanz, C. 1997, "Experimental tasks in SLA Research: Amount of Production, Modality, Memory, and Production Processes," in *Contemporary Perspectives on the Acquisition of Spanish. Volume 2: Production, Processing, and Comprehension*, W. Glass & A. Pérez-Leroux, eds., Cascadilla Press, pp. 41-56.

Sanz, C. & Morgan-Short, K. 2004, "Must computers deliver explicit feedback? An Empirical Study.", in *Processing Instruction: Theory, Research, and Commentary*, B. VanPatten, ed., Lawrence Erlbaum.

Sanz, C. & VanPatten, B. 1998, "On input processing, processing instruction, and the nature of replication tasks: a response to M.Rafael Salaberry.", *The Canadian Modern Language Review*, vol. 54, pp. 263-273.

Schensul, S., Schensul, J., & LeCompte, M. 1999, *Essential ethnographic methods : observations, interviews and questionnaires* AltaMira, Walnut Creek, Calif.

Schmidt, R. 1990, "The role of consciousness in second language learning", *Applied Linguistics*, vol. 11, pp. 129-158.

Schmidt, R. 1993, "Awareness and second language acquisition", *Annual Review of Applied Linguistics*, vol. 13, pp. 206-226.

Schmidt, R. 1994, "Deconstructing consciousness in search of useful definitions for applied linguistics", *AILA Review*, vol. 11, pp. 11-26.

Schmidt, R. 2001, "Attention," in *Cognition and Second Language Instruction*, P. Robinson, ed., Cambridge University Press, pp. 3-32.

Schön, D. 1983, *The Reflective Practitioner: How Professionals Think in Action* Temple Smith, London.

Schwartz, B. 1993, "On explicit and negative data effecting and affecting competence and linguistic behavior", *Studies in Second Language Acquisition*, vol. 15, no. 2, pp. 147-165.

Scott, V. 1989, "An empirical study of explicit and implicit teaching strategies in French", *The Modern Language Journal*, vol. 73, pp. 14-21.

Segalowitz, N. 2000, "Automaticity and attentional skill in fluent performance," in *Perspectives on Fluency*, H. Rigganbach, ed., University of Michigan Press, pp. 200-219.

Segalowitz, N. & Lightbown, P. 1999, "Psycholinguistic Approaches to SLA", *Annual Review of Applied Linguistics*, vol. 19, pp. 43-63.

Sharwood Smith, M. 1981, "Consciousness-raising and the second language learner", *Applied Linguistics*, vol. 2, pp. 159-168.

Sharwood Smith, M. 1986, "Comprehension versus acquisition: two ways of processing input", *Applied Linguistics*, vol. 7, pp. 239-274.

Sharwood Smith, M. 1991, "Speaking to many minds: On the relevance of different types of language information for the L2 learner", *Second Language Research*, vol. 7, no. 2, pp. 118-132.

Sharwood Smith, M. 1993, "Input enhancement in instructed SLA: theoretical bases", *Studies in Second Language Acquisition*, vol. 15, no. 2, pp. 165-180.

Shiffrin, R. & Schneider, W. 1977, "Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory.", *Psychological Review*, vol. 84, pp. 127-190.

Shook, D. 1994, "FL/L2 Reading, Grammatical Information, and the Input-to-Intake Phenomenon", *Applied Language Learning*, vol. 5, no. 2, pp. 57-93.

Siegel, S. 1956, *Nonparametric statistics for the behavioural sciences* McGraw-Hill.

Simard, D. & Wong, W. 2001, "Alertness, orientation, and detection: The conceptualization of attentional functions in SLA.", *Studies in Second Language Acquisition*, vol. 23, pp. 103-124.

Skehan, P. 1996, "A framework for the implementation of task-based instruction", *Applied Linguistics*, vol. 17, no. 1, pp. 38-62.

Skehan, P. 1998, *A Cognitive Approach to Language Learning* Oxford University Press, Oxford.

Slimani, A. 1991, "Evaluation of classroom interaction," in *Evaluating Second Language Education*, C. Alderson & A. Berretta, eds., Cambridge University Press, pp. 197-220.

Slobin, D. 1979, *Psycholinguistics* Scott, Foresman and Company.

Smith, F. 1971, *Understanding Reading* Holt, Rinehart and Winston.

Spada, N. 1997, "Form-focussed instruction and second language acquisition: a review of classroom and laboratory research", *Language Teaching*, vol. 30, pp. 73-87.

Spada, N. & Lightbown, P. 1993, "Instruction and the development of questions in L2 classrooms", *Studies in Second Language Acquisition*, vol. 15, no. 2.

Stenhouse, L. 1975, *An Introduction to Curriculum Research and Development* Heinemann, London.

Stern, H., Burstall, C., & Harley, B. 1975, *French From Age Eight, Or Eleven?* The Ontario Institute for Studies in Education, Ontario.

Swain, M. 1985, "Communicative competence: Some roles of comprehensible input and comprehensible output in its development.", in *Input in Second Language Acquisition*, S. Gass & C. Madden, eds., Newbury House, pp. 235-253.

Swain, M. 1995, "Three functions of output in second language learning," in *Principle and Practice in Applied Linguistics*, G. Cook & B. Seidlhofer, eds., Oxford University Press, pp. 125-144.

Tanaka, Y. 1996, *The comprehension and acquisition of relative clauses by Japanese High School students through formal instruction.*, Ed. D diss, Temple University Japan.

Taylor, G. & Edwards, D. 1994, *Etoiles 3* BBC Books and Longman.

Terrell, T. 1991, "The role of grammar instruction in a communicative approach", *The Modern Language Journal*, vol. 75, pp. 52-63.

Terrell, T., Baycroft, B., & Perrone, C. 1987, "The subjunctive in Spanish interlanguage: accuracy and comprehensibility.", in *Foreign Language Learning: A Research Perspective*, Bill VanPatten, Trisha Dvorak, & James Lee, eds., Newbury House, pp. 19-32.

Todd, L. 1974, *Pidgins and Creoles* Routledge.

Tomlin, R. & Villa, V. 1994, "Attention in cognitive science and second language acquisition", *Studies in Second Language Acquisition*, vol. 16, no. 2, pp. 183-204.

Tooley, J. & Darby, D. 1998, *Education Research: An OfSTED Critique*.

Toth, P. 1997, *Linguistic and pedagogical perspectives on acquiring second-language morpho-syntax: A look at Spanish se*, Ph.D. diss., University of Pittsburgh.

Towell, R. & Dewaele, J. 2002, "The role of psycholinguistic factors in the development of fluency amongst advanced learners of French," in *Focus on French as a Foreign Language: Multidisciplinary Approaches*, J. Dewaele, ed., Multilingual Matters.

Towell, R. & Hawkins, R. 1994, *Approaches to second language acquisition* Multilingual Matters.

Trahey, M. 1996, "Positive Evidence in Second Language Acquisition: Some Long-Term Effects", *Second Language Research*, vol. 12, no. 2, pp. 111-139.

Trahey, M. & White, L. 1993, "Positive evidence and preemption in the second language classroom", *Studies in Second Language Acquisition*, vol. 15, pp. 181-204.

Trim, J. 1988, "Applied linguistics in society", *British Studies in Applied Linguistics*, vol. 3, pp. 3-15.

Truscott, J. 1998, "Noticing in second language acquisition: a critical review", *Second Language Research*, vol. 14, no. 2, pp. 103-135.

Turner, K. 1995, *Listening in a Foreign Language: a skill we take for granted?* Pathfinder 26, CILT.

Tymms, P. 1999, "Homework, common sense, politics and the defence of research", *Research Intelligence*, vol. 70, pp. 22-23.

van Naerssen, M. 1981, *Generalizing second language hypotheses across languages: A test case in Spanish as a second language*, PhD thesis, University of Southern California.

VanPatten, B. 1990, "Attending to form and content in the input: An experiment in consciousness", *Studies in Second Language Acquisition*, vol. 12, pp. 287-301.

VanPatten, B. 1996, *Input Processing and Grammar Instruction in Second Language Acquisition* Ablex, Norwood, N.J.

VanPatten, B. 1997, "The relevance of input processing to second language theory and second language teaching," in *Contemporary Perspectives on the Acquisition of Spanish. Volume 2: Production, Processing, and Comprehension*, W. Glass & A. Pérez-Leroux, eds., Cascadilla Press, pp. 93-108.

VanPatten, B. 2000, "Thirty years of input (or intake, the neglected sibling)," in *Social and Cognitive Factors in Second Language Acquisition*, B. Swierzbin et al., eds., Cascadilla Press, pp. 287-311.

VanPatten, B. 2004, "Input Processing in SLA," in *Processing Instruction: Theory, Research, and Commentary*, B. VanPatten, ed., Lawrence Erlbaum.

VanPatten, B. & Cadierno, T. 1993, "Explicit instruction and input processing", *Studies in Second Language Acquisition*, vol. 15, pp. 225-241.

VanPatten, B. & Oikkenon, S. 1996, "Explanation versus structured input in processing instruction", *Studies in Second Language Acquisition*, vol. 18, no. 4, pp. 495-510.

VanPatten, B. & Sanz, C. 1995, "From input to output:processing instruction and communicative tasks," in *Second Language Acquisition Theory and Pedagogy*, F. Eckman, D. Highland, & et al., eds., pp. 169-185.

VanPatten, B. & Wong, W. "Processing instruction and the faire causatif in French: a replication", Paper presented at the AAAL conference 2001.

Wells, M. 2000, *Reflections on Grammar - implicit language teaching*. CILT.

White, J. 1998, "Getting the learners' attention: A typographical input enhancement study," in *Focus on Form in Classroom Second Language Acquisition*, C. Doughty & J. Williams, eds., Cambridge University Press, pp. 85-113.

White, L. 1991, "Adverb Placement in Second Language Acquisition: Some Effects of Positive and Negative Evidence in the Classroom", *Second Language Research*, vol. 7, no. 2, pp. 133-161.

White, L. 1995, "Input, triggers, and second language acquisition. Can binding be taught?", in *Second language acquisition theory and pedagogy*, F. Eckman et al., eds., Erlbaum, pp. 63-78.

White, R. 1999, "Changes in research since 1969," in *Researching Teaching: Methodologies and Practices of Understanding Pedagogy*, J. Loughran, ed., Falmer Press.

Wickens, C. 1984, "Processing resources in attention," in *Varieties of attention*, R. Parasuraman & D. Davies, eds., Academic Press.

Wickens, C. 1989, "Attention and skilled performance," in *Human Skills*, 2nd edn, D. Holding, ed., John Wiley, pp. 71-105.

Wijers, A., Mulder, G., Gunter, T., & Smid, H. 1996, "Brain potential analysis of selective attention," in *Handbook of Perception and Action Volume 3: Attention*, O. Neumann & A. Sanders, eds., Academic Press, pp. 333-387.

Williams, J. "FonF: What kind of focus? Which form?", Paper presented at Annual AAAL Conference. Long Beach, California, March 1995.

Williams, J. 1999a, "Memory, attention, and inductive learning", *Studies in Second Language Acquisition*, vol. 21, pp. 1-48.

Williams, J. 1999b, "Learner-generated attention to form", *Language Learning*, vol. 49, no. 4, pp. 583-625.

Williams, J. & Evans, J. 1998, "What kind of focus and on which forms?", in *Focus on Form in Classroom Second Language Acquisition*, C. Doughty & J. Williams, eds., Cambridge University Press, pp. 139-155.

Wimpory, M. 1997, *Come back grammar! All is forgiven: Students' & Lecturers' experience of foreign language grammar up to GCSE*, MA dissertation, University of York.

Wong, W. 2001a, "Modality and Attention to Meaning and Form in the Input", *Studies in Second Language Acquisition*, vol. 23, no. 3, pp. 345-368.

Wong, W. Processing Instruction: Examples of Form-Meaning Connections in French.
2001b. Unpublished manuscript

Wong, W. 2002, "Teaching for Form-Meaning Connections in French", *The French Review*.

Woods, A., Fletcher, P., & Hughes, A. 1989, *Statistics in Language Studies* Cambridge University Press.

Wright, M. 1999, "Grammar in the languages classroom: findings from research", *Language Learning Journal*, vol. 19, no. 33, p. 39.

Wringe, C. 1996, "Modern Foreign Languages," in *A Guide to Educational Research*, P. Gordon, ed., Woburn Press, pp. 229-246.

List of Appendices

- 1 Extracts of observation records from action research projects on grammar pedagogy carried out 2000/2001
- 2 Description of current policy initiatives regarding grammar pedagogy
- 3 Example of activity from published material focussing on nouns & semantic processing
- 4 Tabular summary of Processing Instruction studies to date
- 5 Examples of PI materials from published studies
- 6 Review of production of verb inflections amongst learners similar to those in this study
- 7 Graphic models of IP
 - a "An expanded model of Input Processing in SLA", VanPatten 1996
 - b "A detail of the process of input – intake", VanPatten 2004
- 8 Input Processing Principles, VanPatten 2004
- 9 Questionnaire to learners about prior language learning experience
- 10 Record of lesson observations
- 11 Example of an observation record
- 12 Example of a PI unit of materials
- 13 Example of an EI unit of materials
- 14 Record of protocol during experimental intervention
 - a Class A
 - b Class B
- 15 Statistics for analysis comparing PI, EI and class C groups at pre test
- 16 Sequence of linguistic items taught
- 17 Statistics for analysis comparing versions 1 and 2 of the tests
- 18 Language sample in the interpretation tests
- 19 Language sample in the production tests
- 20 Descriptions of the tests
- 21 The interpretation tests (version 1)
 - a Reading tests
 - b Listening tests
- 22 The production tests (version 1)
 - a Writing tests
 - b Speaking tests
- 23 Scoring the written interlanguage

- 24 Scoring the oral interlanguage
- 25 Likert scale questions regarding learners' attitudes to learning French
- 26 Questionnaire to PI and EI learners about intervention materials
- 27 Non-parametric statistical tests
- 29 Tests of the normality of the distribution of the test scores
- 30 Schedule for semi-structured interview with teachers A & B regarding intervention
- 31 Descriptive statistics of the language test scores
- 32 Levene's tests for the equality of variances
- 33 Repeated measures ANOVAs without class C
- 34 Repeated measures ANOVAs of the gains between the tests
- 35 Statistics for analyses of issues surrounding validity and reliability of the writing tests
- 36 Statistics for analyses of issues surrounding validity and reliability of the oral tests
- 37 Selective quotes from learners' responses to questionnaire regarding intervention materials

Appendix 1: Extracts from observation records from grammar pedagogy action research projects

... indicates material has been omitted

Extracts 1 & 2 are from a series of lessons in an action research project focussing on teaching and learning present and perfect tense verb inflections in French.

Extract 1

Time Written in centre of board: Appel: Quand es-tu allé(e) au cinéma?
10:15- ... [T elicits meaning of the question on board]
10:19 ... P: when did you go to the cinema?
T: comment est-ce-qu'on répond? I went P: je aller T: non P: j'ai allée T: non, shhh P: je suis allé T: [to other pupil] when you don't know don't say anything.
T elicits various time adverbs, asking for words that describe when things happen
T: for example last week, last month, ça commence avec un D.
T writes il y a 3 mois T: in French we say 'there are 3 months' for 'ago'.
Writes on board: je suis allé au cinéma la semaine dernière, hier, le mois dernier, il y a 2 mois, 3 mois.
10:20- Ps read a response from board when their name is called out from the register - they
10:27 repeat je suis allé au cinema and replace the 'time adverbial' slot

Extract 2

10:55 blockbusters grid displayed on OHP. T : Fermez vos cahiers, fermez vos livres. Grid has English sentences on e.g. we have watched, last weekend, they watched, I am watching. Etc. Teams must translate as many as possible. Sense of competition acute as team numbers made equal. Encouragement of team members and complaints about unfairness! Teacher checks some vocabulary from grid. Last weekend not known so noted in vocabulary books by some learners. Remember the French will say weekend last. Those who were away can keep books open for reference.
11.05 game begins. Pupils obviously familiar with game. Similar insistence on accuracy - learners must spell endings. Confirmation checks if not accurate. Correction by other pupils on each item, until item correctly produced by a pupil. All pupils must be asked so T suggests they should put hand up early. Most pupils actively trying to work out one of the squares in the grid. T You have to think whether it's in present or past. T asks What page is the vocabulary on? Confusion over ils verb ending.
11.10 game was a draw with 4 squares for each team T explains in order to leave the room, pupils must produce either 'nous avons regardé' or 'nous regardons' + programme, depending on whether teacher says present or past. Pupils line up and give the pattern - allowed to go if correct, if wrong they go to back of line. They generally seem to be smiling and trying to hear what the correct answers are.

Extracts 3 and 4 are from a series of lessons focussing on the teaching and learning of explicit knowledge and ability to use present tense verbs in German.

Extract 3

10.18 Pupils arriving...

Your aims for today (written on board): to be very confident with the ich/du /er-sie/ forms of the verb in speaking and writing.

10.20 T: was machst du in deiner freizeit? P: ich faulenze P: ich gehe ins Kino T: was macht er? P: er faulen T: er faulenzt P: faulenzt und gehe [?] ins kino T repeats. T asks another pupil [...] P: er spielt Klavier, er faulenze [?] T asks another P P: ich gehe schwimmen. Ich lese. T: was machst sie? P: ich lese T: that's what she said P: sie les T: can anyone help? P: liest Next P: ich ... P: sie faulenze und sie liest P: ich tanzt T corrects pronunciation P (reporting in 3rd pers): ich T: ich? Sie P: sie tanze T: but I need it exactly right P – struggling. Another couple of pupils. T: Most of you aren't too bad on that. T: OK, P was machst du gern? Why's that different? P: don't suggest [?]. T: what has gern added to it? 10.27 P: ich gern. T asks another P: why is it different? P: it's different because you're saying what you really like T: try P: ich gern besuche Freunden T: ich besuche gern Freunden P repeats. P: ich gehe gern [?] T: gern P: gern T: I don't want you to confuse this with gehen. Some discipline in L1. P: ich gern [or gehet?] ins Kino

10.32 T: where's the gern? P: reaction shows confusion T: you haven't been listening... Other pupils try P: ich spiele gern... P: ich höre gern Musik T: so what are you telling me about how you do it? If you don't like something?

Extract 4

11.32 Today is a one-off session, to see what [you] know about verbs in English and in German, to give examples and so on. What is a verb? (c6 hands up).
Pupils offer: a doing word, a describing word...an action word...one pupil offers: a doing or a being word...

11.36 T comments on range of examples, says she had not expected e.g.s like having, being, it is good to notice these are also verbs "I am tall"- is there a verb?...

11.39 What about helping verbs, words that turn up near a verb? Ps suggest I, he, she
T says yes, usually a person near the verb...P offers "is" - T says no, that's the verb. P offers 'a', like in 'a person' T says yes....Ps offer "the cat...the thing...a noun...object, place, person, thing..."

11.42 T rotates whiteboard...shows verb paradigms [extract only shown here]:

To drink	trinken	Beve/bevere
I do drink, I am drinking, I drink	Ich trinke	bevo
You do drink, you are drinking, you drink...	Du trinkst...	Bevi...
They do drink, they are drinking, they drink	Sie trinken	Bevono

11.46 T comments: languages are all related...why is there only one word in Italian?...cause they all begin with be-? P asks is it to do with the endings? Is it that "-o" means "I" and so on?...T asks...why does German need pronouns? T: because it doesn't change much...

11.51 T rotates whiteboard again to show pre-planned list of sentences [extract only shown]
1 I do drink wine sometimes
2 He had been drinking that evening
10. She is drinking quickly today
Pupils are asked to pick out verbs; main errors involve producing V with its pronoun, adverbs also cited. T says there is a time idea connected with verbs! What sort of time is connected with [we will drink it all] P Present! No, future! ...
11.58 ...which bit on board [referring to trilingual verb paradigms] has no person? ...P offers, the titles T: these are called infinitive

12.02 T gives out worksheet...you have 10 English sentences, circle verb, not the person, it may be more than one word!

12.09 T says, tell me any V you can think of in German Ps offer lots including: macht, einpacken...bin T asks which endings do you know? If you've got "ich" in German what's the ending? ..for du?, etc ...and if it is er and sie?

12.15 T tells Ps to do the German sentences on the worksheet now, find the Vs in the sentences, but looking just for one word this time [observer's comment: n.b. meaning of sentences is not explained - what clues are the pupils using? Meaning/form? Endings? Pronouns? Intuition?]...at no. 3, a pupil asks what the sentence means, T says oh right, do you want me to tell you what they mean? Some pupils have underlined adjectives...

12.24 Let's check homework, song was on p45, you were trying to find me the verbs!...T reads out the song phrase by phrase, asks for words...

Appendix 2: Government initiatives and their implications for grammar pedagogy

Specialist Language Colleges

Specialist Language Colleges (currently about 200 in the UK) may be able to provide guidance, certainly at a local level, in terms of grammar instruction, in that they are perhaps more likely to engage in, host and provide professional development activities. It is a condition of their status as Specialist schools that learning at least one language at GCSE is obligatory.

The (Revised) National Curriculum (NC) (DfEE 1999)

The 1999 MFL NC now suggests that there is an interface between knowledge about language and the ability to use language:

“During Key Stage 3... pupils become familiar with the sounds, written form and grammar of the language, and use this knowledge with increasing confidence and competence to express themselves in role plays, conversation and writing”.

In addition, it is suggested that “identifying the grammatical function of unfamiliar words” (Programme of Study, 3b) will help learners to “interpret meaning”. However, the levels of achievement (from 1 to 9) in reading, listening, speaking and writing provide only a very broad indication of the grammar learners 'should' be able to comprehend and produce and when (further discussion of the validity or usefulness of the NC is given in section 1.3).

Schemes of Work (QCA 2000a)

The non-statutory guidance provided by the KS 3 Schemes of Work (SoW) provides a clear indication of what is thought to be good practice for systematically incorporating specific language structures into a functional syllabus, asserting that some focus on form is beneficial (including recommendation of specific error correction strategies) and that metalinguistic knowledge will help some learners.

Criteria for GCSE (QCA 2000b)

The GCSE specifications contain a list of grammatical structures which could be included in the examinations. 20% of the total GCSE grade is now explicitly allocated to “knowledge and accurate application of the grammar and structures”, awarded in the

writing and speaking exams (see Marsden 2001 for the argument that this has not substantially changed from previous years, where accuracy has always been a high priority). All 3 examination boards in their sample GCSE written papers for 2003 include a 'verb paradigm' activity requiring learners to fill in gaps with accurately inflected verbs, reminiscent of grammar translation or behaviourist style exercises. Ample anecdotal evidence, personal experience and empirical studies (Chambers & Richards 1993) show that there is considerable 'backwash' from these exams (for example, at a training day for teachers regarding the new GCSE syllabus one of the awarding body's representatives advised attendees to start teaching the GCSE content and using their assessment criteria at KS 3). For example, publishers of new text books have been keen to incorporate practice relevant to the new GCSE exams {CILT 2001 474 /id}, including more grammar explanations and some verb paradigm activities.

Key Stage 3 Framework

The non-statutory 'Key Stage 3 Framework for MFLs' reflects a belief that MFL teaching can be usefully modelled along the lines of the National Literacy Strategy for English {DfEE 1998 465 /id} and {DfEE 2001 54 /id}. There is an emphasis on developing learners' metalinguistic and explicit knowledge about language and a heavy use of the word, sentence, text level approach of the NLS.

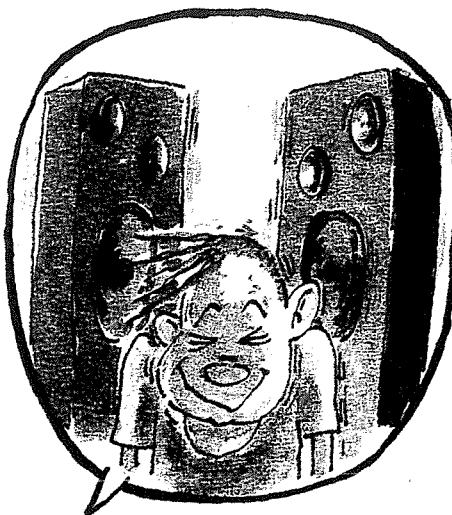
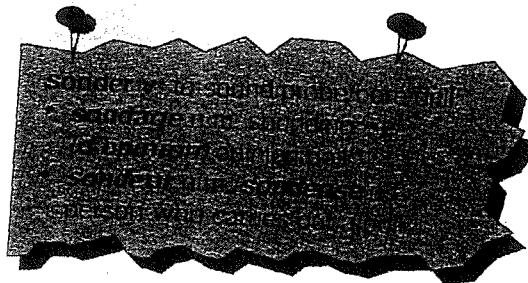
Appendix 3a Example of listening and reading activities in popular text book
(McNab 1994)

Un sondage

⚠ 1 Fermez le livre ou cachez la page!
Combien de questions est-ce que vous pouvez formuler en deux minutes?

Exemples:

Aimes-tu écouter de la musique pop?
Où habitez-vous?



2 a Ecoute: Tutoyer ou vouvoyer? (1-10)
Mets un T s'ils utilisent la forme 'tu' et un V s'ils utilisent la forme 'vous'.

On parle avec un copain – on se tutoie.

Exemple:

Où habites-tu?

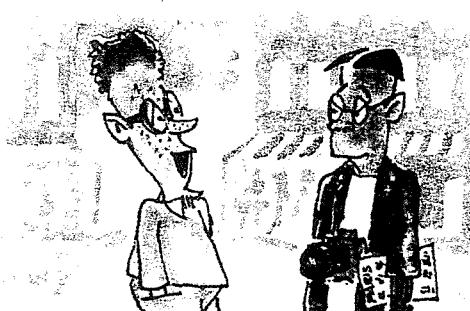


tutoyer = *to use the tu form*
vouvoyer = *to use the vous form*

On parle avec un adulte – on le vouvoie.

Exemple:

Où habitez-vous?



b Ils tutoient ou vouvoient? (T ou V)
Ecris la forme qui manque.

- 1 Où est-ce que tu habites?
- 2 Qu'est-ce que tu portes?
- 3 Préférez-vous le café ou le thé?
- 4 As-tu des frères ou des soeurs?
- 5 Avez-vous des ciseaux?

- 6 C'est quand votre anniversaire?
- 7 Vous êtes de quel signe?
- 8 Aimes-tu regarder la télé?
- 9 Vous partez à quelle heure?
- 10 Est-ce que vous avez déjà fait du surf?

Appendix 3b Example of listening and reading activities in popular text book (McNab 1994)

A VOS MARQUES!

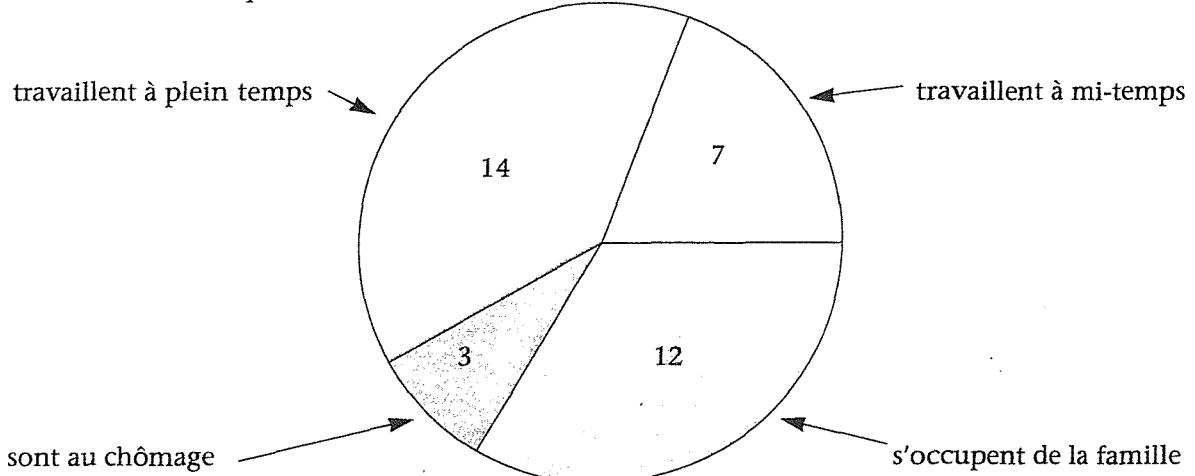
3 Que font nos parents?

a Ecoute: Les résultats du sondage de Nicolas. Copie et complète le texte.

Il y a pères qui travaillent, qui sont au chômage et qui ne travaille(nt) pas.
Parmi ceux qui travaillent:
..... % travaillent dans un bureau
..... % travaillent dans une usine ou un atelier
..... % travaillent à leur compte
..... % etc.

b Regarde le graphique et écris une conclusion.

Nombre de mères qui:



Fais un sondage.

a Choisis un thème.

Exemples:

la famille; la maison; ma chambre; les loisirs etc.

b Prépare une/des question(s).

Exemples:

As-tu des frères ou des soeurs?

Aimes-tu ...?

Préfères-tu ...?

Comment trouves-tu ...? Bon, bof, nul?

c Pose la/les question(s) à douze/vingt-quatre copains et copines.

d Dessine un graphique et écris tes conclusions.

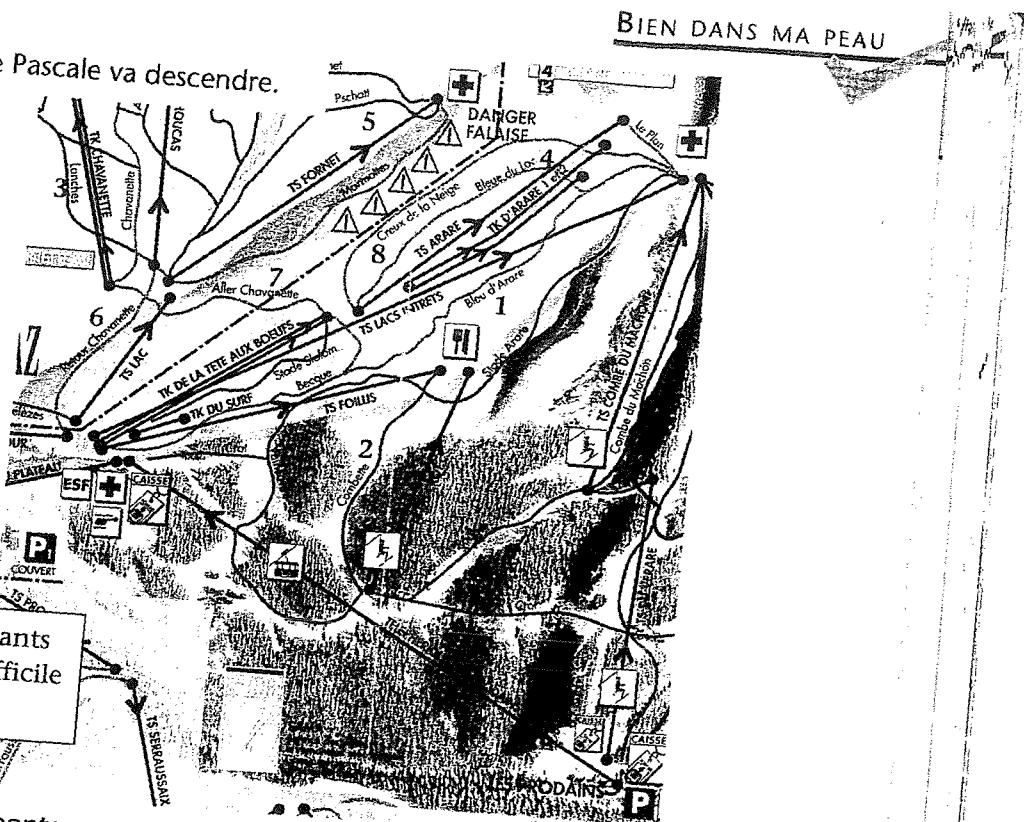
à toi

pare et enregistre les conclusions de ton sondage.



Appendix 3c Example of listening and reading activities in popular text book (McNab 1994)

Ecoute; Trouve les pistes que Pascale va descendre.



a. Ecoute et lis: Une journée à la montagne

Quelle journée fatigante ! Je me suis levée à 7h, j'ai avalé mon petit déjeuner et nous sommes partis en voiture.

Le trajet a duré une heure. A 9h nous étions sur les pistes. Nous avons profité des grandes descentes toutes blanches, car il n'y avait encore personne.

Cela faisait deux ans que je n'étais pas montée sur des skis, mais je me suis quand même bien débrouillée.

Evidemment, je suis tombée quelques fois au début.

En tout cas, le ski, ça donne faim ! A midi, j'ai dû avaler 5 ou 6 sandwichs et boire un demi-litre de jus d'orange. J'étais avec mon père et mon oncle.

Enfin, on a bien rigolé et je me suis bien dépensée, c'est le principal, non ?

Chez toi

Qu'est-ce qu'ils ont fait?

Exemple:

Exemple.

b Vrai ou faux?

- 1 Pascale est allée au ski avec ses parents.
- 2 Ils sont partis à neuf heures.
- 3 Ils y sont allés en voiture.
- 4 Le voyage a duré une heure.
- 5 Il y avait beaucoup de monde sur les pistes.
- 6 Il faisait du brouillard.
- 7 Elle est beaucoup tombée.
- 8 A midi elle avait très faim.
- 9 Elle avait très soif.
- 10 Elle s'est bien amusée.

c Corrige les phrases qui sont fausses.



Appendix 4 - Summaries of Processing Instruction Studies to date

Abbreviations used (providing definitions for these terms was beyond the scope of this summary)

PI Processing Instruction (as defined in VanPatten & Cadierno 1993)

TI Traditional instruction (fairly mechanical output practice inc. whole paradigms)

DPT Delayed post test

EGI Explicit grammar instruction

MOI Meaning oriented instruction (production)

OP Output practice

Sig. Statistically significant

Some information was not found in the published studies

Studies with findings largely supporting the role of PI (compared to 'output practice)

<i>Study; Target language forms</i>	<i>L1; Educational level; Instructional Groups (n); Length of instruction</i>	<i>Assessment tasks; Timing of post & DPT after intervention</i>	<i>Results & comments</i>
VanPatten & Cadierno 1993 Morphosyntax of Spanish object pronouns	2 nd year univ; English; PI (27), TI (26), Control – read & discussed essay (27) 2 hours over 2 days	Aural interpretation sentence: 10 target items and 10 distracters; Written production, sentence: five items with five distracters; Immediate, 2 weeks & 1 month	Interpretation: PI – sig. gains maintained on DPT, TI & Control - no gains Production: both PI and TI made and maintained sig. gains at DPT, Control – no gains
Cadierno 1995 Spanish preterit tense	2 nd year college; English; PI, TI, Control	Aural interpretation (20 items) Written production (5 items sentence completion); Immediate, 2 weeks & 1 month	Interpretation: PI gains & maintained on DPTs TI & Control did not improve and not different from each other Production: PI & TI improved and maintained improvement. No difference between PI & TI. Control - no gains

Cheng 1995 & 2002 Spanish copulars ser & estar	2 nd year college; English; PI, TI, Control	Interpretation & Production (Written composition based on pictures as in VP & C 1993) Tests focused on use of <i>estar</i> , as <i>ser</i> is learners' default; Immediate & 3 weeks	Interpretation: PI and TI improved. PI improved more. On DPT PI dropped to same level as TI. Production = PI and TI improved and both retained gains on DPT. Written composition = both made gains and both retained gains on DPT. Control group made no gains
Farley 2001 Spanish subjunctive with noun clauses and expressions of doubt and negation	English - Spanish PI MOI 2 days MOI = no mechanical drills and based on 'structured-output activities' from Lee & VanPatten (1995), participants created subordinate clauses using subjunctive or indicative forms based on the main clause triggers heard (or read). Affective activities = expressed own beliefs/ doubts	Aural interpretation 12 target items (listening to verbs to choose suitable written clauses); Written production 12 target items; Immediate and DPT one month	Interpretation = significant gains Production = significant gains No difference between groups
Farley 2001b As above	As above		Both groups improved on both interpretation and production, only the PI group maintained its performance on delayed task. MOI group declined in performance.
Buck 2000 English present continuous (vs. the present progressive)	University Spanish; PI, TI	Interpretation: e.g. <i>Bill is smoking a pipe</i> Vs <i>Bill smokes a pipe</i> Production: e.g. "I ____ to music every day [listen]" and "We ____ the new rota this week [prepare]").	Interpretation: PI better than TI, maintained over time; TI gains not maintained. Production: PI & TI made similar gains and maintained over time.

VanPatten & Wong 2001 French causative	English; University; PI, TI, Control	Aural Interpretation 14 sentences Written Production 10 sentences	Interpretation: PI better than TI; Control = no improvement. Production = PI & TI improved significantly but were not different from each other. DPT: interpretation PI did not maintain improvement (due to absenteeism)
Benati 2001 Italian future tense	English; University PI, TI (explanation of grammar rules followed by written and oral practice, part of which was 'meaning oriented'), Control (39 total)	20 aural interpretation sentences; 2 part production task – 5 written sentence-level and 5 oral limited response from pictures; Pretest, immediate Posttest; DPT - 3 weeks	Interpretation: PI & TI improved significantly; Control = no improvement. PI gains significantly greater than TI (PI > TI > C). Both Production tasks = PI and TI groups both improved with no difference between them; control did not improve. (improvement of TI on interpretation possibly due to minimized use of mechanical form-oriented activities)
VanPatten & Sanz 1995. Morphosyntax of Spanish object pronouns	English; 2 nd year univ; PI, Control ('no instruction'); 2 hours	interpretation test from VanPatten & Cadierno 1993 3 kinds of output tests all in both oral and written modes: sentence-level test 14 items, question-answer test based on pictures approx 11 items and video-narration test; Immediate post test	Control = no improve on any tests. PI = improved significantly on interpretation test and on the sentence-level test in both oral & written. Question-answer test = significant gains but not many object pronouns produced Video-narration = gains were significant in written mode but 'just missed' significance in the oral mode PI performed better on written than oral.

Studies related to exploring the role of EGI in PI

<i>Study; Target language forms</i>	<i>Educational level; L1; Instructional Groups (n); Length of instruction</i>	<i>Assessment tasks; Timing of post & DPT after intervention</i>	<i>Results & comments</i>
VanPatten & Oikkenon 1996 Morphosyntax of Spanish object pronouns	4 th semester high school students; English; PI (17), EGI only (22), Referential + affective only (20); '4 days'	10 target items in interpretation, 5 items in production test (as in VanPatten & Cadierno) immediate post tests	Production: the EGI + ref&aff group better than the EGI-only group. EGI and practice-only same (i.e. the <i>combination</i> of the EGI with the practice which makes sig. diff. (their mean scores went up to 3 out of 10 compared with 2.15 and .78 in and control)
Benati 2004 Regular Italian future tense	2 nd semester University students English PI (14), Referential & affective only (12), EGI only (12). Taught twice (3 hours each) over two consecutive days	interpretation 20 aural sentences (10 present "as distractors" and 10 future). written production task short text with five blanks for future tense verbs Immediate post test DPT one month	The PI and the ref + affect only groups improved significantly more than the EGI only. NO difference between PI and ref & affect only.
Sanz & Morgan-Short 2001 Morphosyntax of Spanish object pronouns	University; English; 4 groups all with structured input but: 1 Explicit feedback only; 2 Explanation only; 3 Explicit feedback + explanation; 4 No explicit feedback no explanation.		No significant difference between any of the groups – neither explicit feedback nor explanation, nor combination had differential effect. Video narration – significant gains

Studies which did not satisfactorily operationalise PI (both according to VanPatten and supported here)

<i>Study; Target language forms</i>	<i>Educational level; L1; Instructional Groups (n); Length of instruction</i>	<i>Assessment tasks; Timing of post & DPT after intervention</i>	<i>Results & comments</i>
DeKeyser & Sokalski 1996. Morphosyntax of Spanish object pronouns & Conditional	University; English; EGI the same for both groups and kept available during tasks Altered PI (input practice), Altered TI (output practice), Control, Kept “the need for meaning as similar as possible for the two treatment groups” (p. 625). Followed the progression of mechanical to meaningful to communicative exercises for <u>both</u> experimental treatments (p. 626). Did not control for the effect of teacher. Small sample?	Immediate, DPT 3? weeks	<p><i>Object pronouns</i> Comprehension = input better. Production = output better Control, input & output similar on DPT.</p> <p><i>Conditional</i> Comprehension: output better, but not on DPT Production: output better, but not on DPT. Input did not drop in performance from the immediate to DPT; the output group did Pretests showed that subjects seemed more proficient in target structures than subjects in other studies</p>
Salaberry 1997 Spanish object clitic pronouns	3 rd semester university; English “IP”, ‘Output Processing’, Control (average 10 in each group); One and a half hours of instruction.	written comprehension 10 items (v similar to an input instructional activities); Production: written translation of 6 sentences with object pronouns and written narrative production based on 1 min silent video. DPT interpretation only one month	Both experimental groups improved significantly on all measures with no difference between them. free oral narration test but produced few obligatory occasions for the target structure Did not control for the teacher variable.

Studies not reviewed directly (see Ellis 1999)

<i>Study; Target language forms</i>	<i>LI; Educational level; Instructional Groups (n); Length of instruction</i>	<i>Assessment tasks; Timing of post & DPT after intervention</i>	<i>Results & comments</i>
Tuz 1992 English word order	Japanese; university; Presentation for both groups=16 OHP pictures and verbs listened to & then pairs of pictures illustrated difference in meaning given by word order PI (not authentic according to VanPatten, personal communication): "practice comprehension of sentences" OP same pairs of pictures as stimuli for production practice.	Reading comp; written production - both based on pictures	"Strikingly in favour of the processing instruction group on both comprehension and production post tests".
Tanaka 1996 English relative clauses	high school students; Japanese ; PI, OP(total n=123) both groups had explicit instruction (not known whether this was EGI as in PI).	aural comprehension; written production post tests 5 days, DPT 2 months	Comprehension: PI outperformed OP Production: both improved; OP outperformed PI on immediate post test but not on DPT
Toth 1997; reflexive pronoun <i>se</i> in Spanish	PI Output a) = task-based Output b) = question and answer routines, explicit grammar explanations to all groups (not known whether EGI as in PI).	grammaticality judgement test to measure comprehension	Both types of instruction resulted in gains in accuracy but OP led to more frequent use of the target structure.

Studies which have not had results supporting the benefits of PI with various comparison instructions

<i>Study; Target language forms</i>	<i>L1; Educational level; Instructional Groups (n); Length of instruction</i>	<i>Assessment tasks; Timing of post & DPT after intervention</i>	<i>Results & comments</i>
Collentine 1998; Spanish subjunctive in adjectival clauses involving indefinite antecedents	<p>Explicit phase same for both “before instructional treatments, learners instructed in the form of the subjunctive (not its use) so that morphology of the subjunctive would not be an issue”.</p> <p>PI: matching subjunctive and indicative sentences in Spanish to correct situations or pictures; responding to sentences containing either subjunctive or indicative verb phrases. ‘Heavy’ referential activities, no affective activities.</p> <p>OP completed fill-in-the-blanks in pairs learners had to construct sentences to describe something and appropriately select the subjunctive or indicative as they formulate their sentences.</p> <p>Control</p>	<p>Both aural and written interpretation (10 sentences each), validating sentences against a picture.</p> <p>Written production: fill-in-the-blank.</p> <p>Posttest a day after treatment</p>	<p>Both groups improved significantly compared to control (1.5 to 4.3/10).</p> <p>No difference between PI and OP.</p>
Allen 2000; French causative faire	<p>“PI” (not in opinion of VanPatten), TI, Control</p>	<p>Aural interpretation 15 sentences (“select the picture that goes with what you hear”)</p> <p>Written production 5 sentences (“write 5 sentences about what your parents make you do”).</p>	<p>Interpretation = both groups made significant gains, no differences from each other;</p> <p>Production = both made significant gains, TI more than PI</p>

Appendix 5 Example of PI materials from VanPatten & Wong 2001, teaching ‘causative faire’ to undergraduate students

Explicit grammar instruction

We often ask or get people to do things for us by telling them to do something.

Paul says, “John, would you mind doing the dishes?”

If you and I were to describe what is happening we might say:

We say, “Paul gets John to do the dishes.”

or

“Paul makes John do the dishes.”

This is called a causative construction (because someone is causing a behavior in someone else.) French has a similar structure using the verb *faire*. Let’s repeat our examples from above.

Paul says, “Jean, pourrais-tu faire la vaisselle?”

We say, “Paul fait faire la vaisselle à Jean.”

How would we describe the following scenario?

Wynne says, “Sara, pourrais-tu promener le chien?”

We would describe Wynne getting Sara to do it like this.

We say, “Wynne fait promener le chien à Sara.”

Often we don’t mention who we get to do something; we might simply say we have something done.

“Paul fait nettoyer la chambre.”

In this case, Paul has the room cleaned, but we don’t know who or how.

One of the problems the *faire causatif* presents is in listening comprehension. Second language learners of French often misinterpret what they hear because the word order is different from English. For example, it is not uncommon for learners of French to make the following mistake:

They hear: “Jean fait faire la vaisselle à Paul.”

They incorrectly think: John is doing the dishes for Paul.

or

They hear: “Marc fait couper les cheveux.

The incorrectly think: Marc cuts hair.

In the activities that follow, we will practice hearing and interpreting the *faire causatif*.

Aural Referential Activity

Listen to each sentence, then indicate who is performing the action by answering each question.

1. Who cleans the room? _____
2. Who packs the bags? _____
3. Who watches the movie? _____
4. Who plays the flute? _____
5. Who does the dishes? _____
6. Who buys wine? _____
7. Who watches the show? _____
8. Who reads the instructions? _____

Teacher's script:

Read each sentence ONCE. After each sentence, ask for an answer. Do not wait until the end to review answers. Students do not repeat or otherwise produce the structure.

1. Claude fait nettoyer la chambre à Richard.
2. Marc fait les valises pour Jean.
3. Sandra fait voir le film à Pierre.
4. Louis fait jouer de la flute à Suzanne.
5. Georges fait la vaisselle pour Louis.
6. Louise fait acheter du vin à Diane.
7. Ma mère fait regarder le spectacle à mon père.
8. Sally fait lire les instructions à Jean Luc.

Aural Affective Activity

In this activity you will hear a series of sentences about what a university does and does not make a typical student do. Indicate whether you like the obligation or not.

Ça me plaît

Ça ne me plaît pas

1.	<input type="checkbox"/>	<input type="checkbox"/>
2.	<input type="checkbox"/>	<input type="checkbox"/>
3.	<input type="checkbox"/>	<input type="checkbox"/>
4.	<input type="checkbox"/>	<input type="checkbox"/>
5.	<input type="checkbox"/>	<input type="checkbox"/>

How many do you and your classmates have the same? Which items did you all indicate that you liked? That you didn't like?

Teacher's script:

Read each statement once. Repeat only if a student asks for it. After reading all statements, go back and review with the class to see how they responded. Follow examples from previous activities. Students do not repeat or otherwise produce the structure.

1. L'université fait suivre un cours de langue étrangère à l'étudiant.
2. L'université ne fait pas suivre un cours de philosophie à l'étudiant.
3. L'université fait décider la carrière à l'étudiant trop tôt.
4. L'université fait terminer les études dans quatre ans à l'étudiant.
5. L'université ne fait pas manger à la cafétéria à l'étudiant.

Appendix 6 Evidence regarding the emergence and learning of French inflectional verb morphology

Researchers working with naturalistic and immersion and non-immersion formal learners at various levels (Dietrich, Klein & Noyau 1995, Bardovi-Harlig 2000, Sato 1990, von Stutterheim 1986 & 1991 (in Mitchell & Myles 1998), Howard 1999, Macrory & Stone 1996 & 2000, Labeau 2002, Mitchell & Dickson 1997) have found that learners use a range of non-morphological means to express person, number and tense, e.g. pragmatic (narrative structure) and lexical (temporal adverbials, inherent semantic verb features), and that morphological means are relatively late to develop, particularly in semi-spontaneous oral production.

This appendix presents evidence from a range of synchronic and diachronic studies, both cross-sectional and longitudinal, that have documented instructed French L2 learners' acquisition of verb inflections in the present and perfect tenses, mainly in oral production but also in writing. The learners used in the studies reviewed are in the same education system (i.e. English secondary) as those in the current study, though the learners' hours of classroom exposure to instruction ranges from about 80 to 800 (the year 9 learners in this study had approximately 182 hours). This broad band of educational stages is considered necessary, as judging the 'emergence' or 'acquisition' of forms at one particular stage is difficult for several reasons e.g.: individual learner variation makes cross-sectional comparisons hazy; a form may be 'emergent', or susceptible to instruction to render it emergent, even though the learners does not produce it; backsliding is likely to occur.

a) *Emergence of inflection for person and number in the present tense in semi-spontaneous oral production*

Data from Mitchell & Dickson (1997) suggested that learners first use mainly nouns but by year 8 many begin to use non-finite verb forms. By year 9 a few learners start to inflect for person and number outside routinised formulas, also supported by data collected for the Linguistic Development project¹. Macrory & Stone (1996) also showed that in years 10 and 11 some learners produced some target-like inflections. Page (1999) found that learners with about 624 hours of exposure to instruction usually produced present tense verbs in the finite form.

However, the studies reviewed did not elicit a significant number of obligatory contexts of present tense 1st person plural inflections, even though these can be a key indicator of inflection (as a learner's null ending could constitute an uninflected 'short' form and an apparent 2nd person plural 'ez' may be a non-finite form).

b) Emergence of verb inflections for perfect tense in semi-spontaneous oral production

Bardovi-Harlig's (2000) meta-research found that the perfect tense seems to be one of the first to be marked morphologically to differentiate it from the present, regardless of L1/L2 language pairs.

The over-application of the 1st conjugation paradigm in the formation of the past participle was noted in all the studies reviewed e.g. finiss[e] connaiss[e], ri[e], voul[e], buv[e], ven[e] (Page 1999, Macrory & Stone 2000, Harris 1988).

Page (1999), Labeau (2002) and Howard (1999) (the latter two with learners with 800+ hours exposure) and Harley (1992) suggest that it is unlikely that past participles of the 2nd and 3rd conjugations are emergent amongst year 9 learners as these forms can still be non-target-like with more advanced learners.

Myles (2002) and Rule & Marsden (2004) found that use of auxiliaries was emerging amongst some learners in years 9, 10 and 11, though it was marked by variation and highly idiosyncratic traits. In particular, there was considerable variability within individual learners' use of the third person auxiliaries. It has been frequently noted (Harley 1992, Macrory & Stone 1996 & 2000, Harris 1988, Page 1999) that learners over-generalise the auxiliary *α* in the early stages², and indeed all forms of *avoir* (to verbs that take *être*). Page (1999) and Macrory (1996) suggested that learners were more likely to assign the auxiliary according to the subject, rather than the lexical verb, and this was

¹ ESRC project no. R000223421 directed by Florence Myles, see <http://www.regard.ac.uk>.

² Grégoire (1947 in Clarke 1985) found that most [subject/verb] agreement errors seem to involve a singular verb used with a plural subject. Indeed, in Page 1999 the learners' use of 'α' in 3rd person plural contexts was double that of 'ont' in 3rd person singular contexts, though this error was never seen in the first person e.g. *je sommes, *nous ai.

particularly the case with first person singular verbs (i.e. learners had a preference for either 'j'ai' or 'je suis' regardless of the lexical verb).

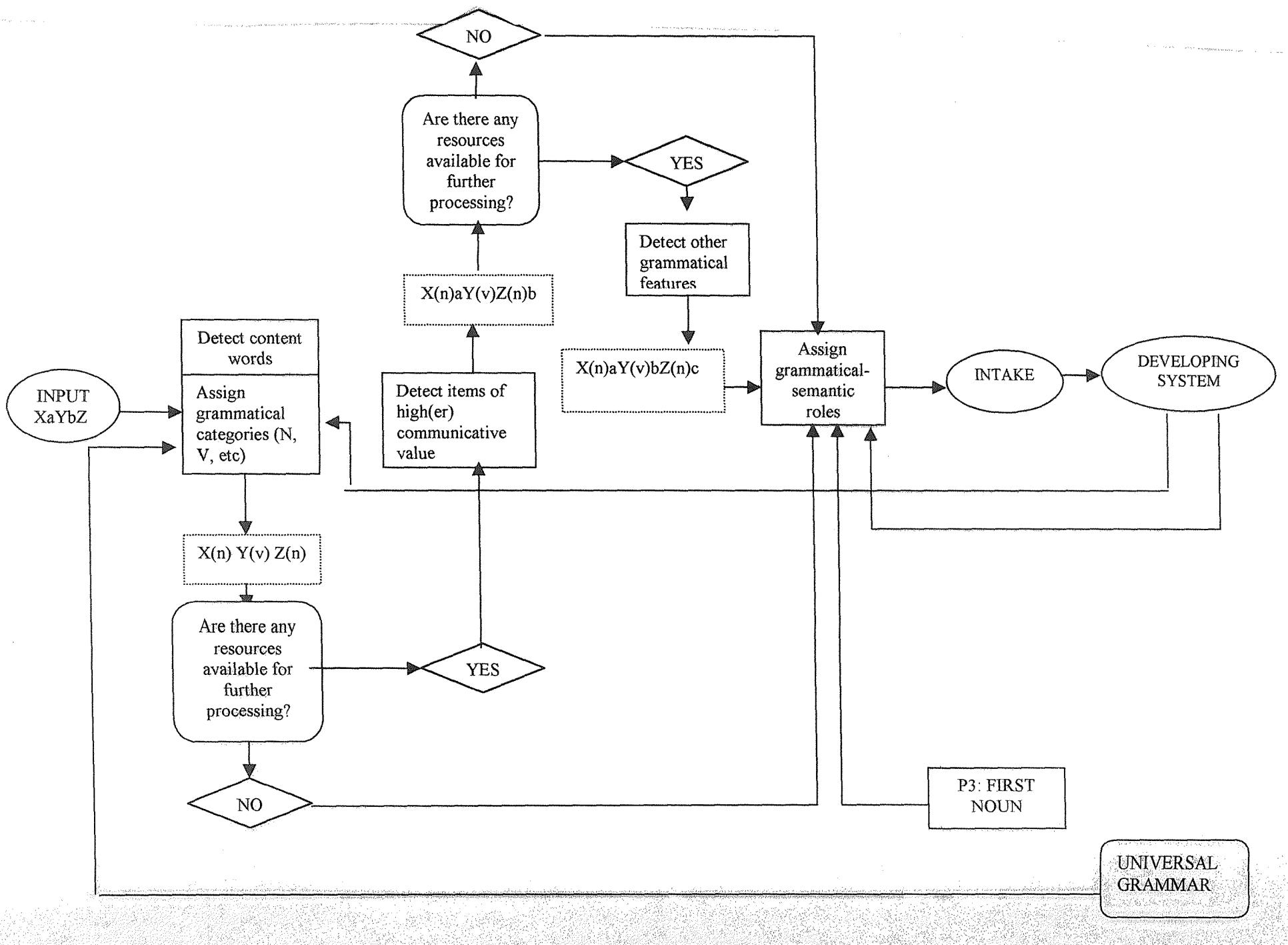
c) *Emergence of verbal inflection for person and number in present and perfect tenses in writing*

This literature search located only four main sources for tracking the emergence of verb inflection in writing: Macrory & Stone (1996 & 2000), Page (1999) and an action research report (which cannot be referenced to maintain anonymity). In addition, school inspectors' reports, DfEE examples of NC levels, pilots for this study and personal experience provide evidence that very early learners can produce a range of written verb inflections for person, number and tense, whose accuracy can depend on task demands (e.g. plannedness, level of production – word, sentence or discourse and content - abstract or concrete). Most sources reviewed note that learners' written production of verbal inflections tended to be more target-like than their oral production.

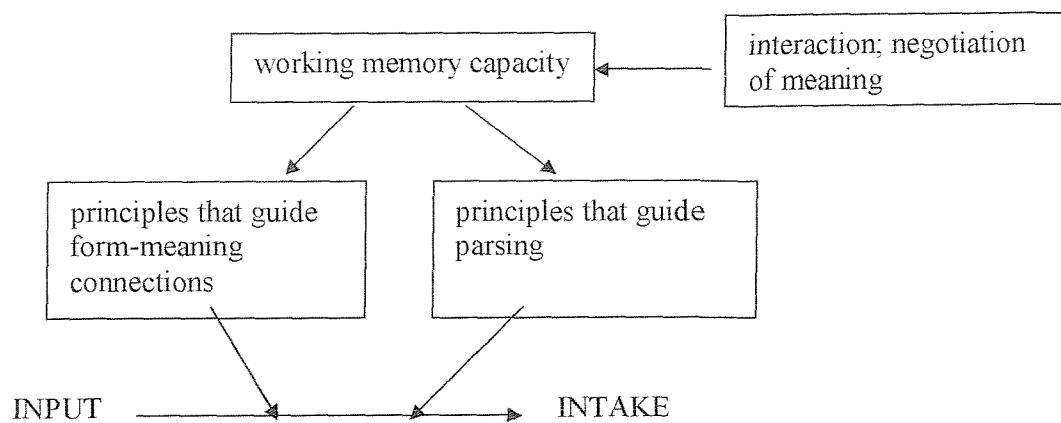
d) *Relationship with the emergence of the pronominal system*

That verb inflections for person, number and tense in French are acquired in contingency with subject clitics is well-documented in L1 acquisition and gradually becoming so in L2 (Prévost & White 2000, Myles 2003 and Rule & Marsden 2004, the latter two regarding similar learners to those in the current study). If this is the case, then it may be futile to attempt to instruct learners in inflectional verb morphology if the pronominal system was clearly non-emergent. Myles, Mitchell & Hooper (1999), Myles (2003), the action research project mentioned above with year 9 learners and pilots for this study found that subject pronouns are productively emergent by years 8 and 9. Rule & Marsden (2004) found that a few learners in year 9 and more in years 10 and 11 were using subject pronouns which behaved as clitics. The action research project also suggested that year 9 learners are beginning to *interpret* subject pronouns correctly in both oral and written input. However, data regarding the developing pronominal reference system was not analysed in this study, partly because most of the elicitation prompts required a new subject (requiring nominal subjects), or the pronouns were provided for the learners.

Appendix 7a Graphic models of IP: An Integrated model of second language input processing for grammatical form and role assignment (VanPatten 1996, p43)



Appendix 7b Graphic models of IP: A detail of the process Input – Intake
[see Figure 1.4, chapter 1]
(VanPatten 2004)



Appendix 8 IP principles (VanPatten 2004)

Principle 1. The Primacy of Meaning Principle. Learners process input for meaning before they process it for form.

Principle 1a. The Primacy of Content Words Principle. Learners process content words in the input before anything else.

Principle 1b. The Lexical Preference Principle. Learners will tend to rely on lexical items as opposed to grammatical form to get meaning when both encode the same semantic information.

Principle 1c. The Preference for Non-redundancy Principle. Learners are more likely to process non-redundant meaningful grammatical form before they process redundant meaningful forms.

Principle 1d. The Meaning-before-nonmeaning Principle. Learners are more likely to process meaningful grammatical forms before nonmeaningful forms irrespective of redundancy.

Principle 1e. The Availability of Resources Principle. For learners to process either redundant meaningful grammatical forms or nonmeaningful forms, the processing of overall sentential meaning must not drain available processing resources.

Principle 1f. The Sentence Location Principle. Learners tend to process items in sentence initial position before those in final position and those in medial position.

P2. The First Noun Principle. Learners tend to process the first noun or pronoun they encounter in a sentence as the subject/agent.

Principle 2a. The Lexical Semantics Principle. Learners may rely on lexical semantics, where possible, instead of word order to interpret sentences.

Principle 2b. The Event Probabilities Principle. Learners may rely on event probabilities, where possible, instead of word order to interpret sentences.

Principle 2c. The Contextual Constraint Principle. Learners may rely less on the First Noun Principle if preceding context constrains the possible interpretation of a clause or sentence.

Appendix 9: Questionnaire about language learning experience outside school

Name: _____

1) Do you have any contact with French speaking people (apart from at school)? Yes / No

If yes: Who is it? _____

How often do you talk/write to them in French? _____

For how long do you normally talk to them in French?

2) Did you learn French before coming to secondary school? Yes / No

If yes: for how long? _____

3) Who was your French teacher in year 7 _____

in year 8 _____

4) Have you ever been to France? Yes / No

If yes: when? _____ (months & year)

For how long? _____

Did you hear much French? a few sentences in total / a bit every day / quite a lot (underline)

Did you speak French? a few words in total / a few words every day / quite a lot (underline)

5) Do you learn or speak any other languages? Yes / No

Which? _____

[-----Following questions only before post and dp tests-----]

• Since Christmas, have you practised French outside school any more than usual? Y

/ N If yes, who with? _____ for how long?: _____

• If you have been to France since Christmas this year:

How long was it for? _____

Who did you stay with? penfriend / someone else: _____

• Are you studying French in year 10? Y / N

If not, please could you explain why not? _____

Thank you !

Appendix 10: Record of data collected during non-experimental intervention

A date given alone indicates observation notes & audio / video recordings available
 report = self report of teacher, prior to or after lesson during informal discussion
 lp = lesson plans collected

Class A	Class B	Class C
17-Sep	21-Sep	lp - 1 Oct
05-Oct	04-Oct	lp - 3 Oct
19-Oct	16-Oct	lp - 15 Oct
22-Oct	18-Oct	lp - 24 Oct
24-Oct	23-Oct	lp - 5 Nov
09-Nov	06-Nov	lp - 7 Nov
14-Nov	13-Nov	obs, audio & lp - 14 Nov
23-Nov	22-Nov	lp - 19 Nov
26-Nov	27-Nov	lp - 21 Nov
12-Dec	07-Dec	lp - 28 Nov
17-Dec	18-Dec	obs & audio 3 Dec
		obs & audio 12 Dec
		17-Dec
PRETESTS	PRETESTS	PRETESTS
see record of intervention	see record of intervention	lp - 23 Jan
		lp - 28 Jan
		lp - 30 Jan
		5 Feb - report
		obs, lp, audio & video 6 Feb
		lp - 18 Feb
		lp - 20 Feb
		? Feb - report
		obs, lp & video 4 March
		lp - 6 March
POST TESTS	POST TESTS	POST TESTS
10-Apr	16-Apr	12 April - report
? Apr - report	18 Apr - pupil report	17-Apr
24-Apr	23-Apr	24-Apr
29-Apr	30-Apr	29-Apr
13-May	16-May	3 May - report
? Apr - report	? May - report	13-May
? Apr - report	? May - report	? May - report
27-May	? May - report	27-May
	28-May	31-May
	? May - report	
DPT	DPT	DPT

Appendix 11 Example of an observation record for this study

Year 9 French
Teacher: Teacher A
Observer: EM

School A
Monday 17 September 2001
10:11:10

B B	G G
G G G G	G G G G
B B B B	G G G G
	B B B
Obs G	B B B

12 boys
15 girls

10.12 Teacher A keeps pupils in same place from lesson to lesson, comments to me whilst giving out books this helps her to remember who they are and faces. Teacher A shows she knows where absent pupils should be, handing out books and leaving them on the desk in their place for when they arrive
Aims written on board: Continue work in all 4 skill areas, (L,S, R, W) on giving details for he/she
Look at m. and f. job names

T checks who is away – knows who is away and asks pupils where they are, making a note in register
T introduces me – stressing to pupils not to take things personally if I come and listen to them and the things I do say to them. They will not have a moment of fame, your names will not appear anywhere, my name won't even appear anywhere, we are not allowed to mention names, the name of the school can just about get mentioned. We are interested in what's going on in your heads, not trying to judge whether you did or said things wrong, we don't mind if you make mistakes (EM adds we like that! we like it when you make mistakes, that's interesting) Teacher A continues, it's all interesting, it's fascinating, we're trying to work out, what's going on when you decide to do certain things or try to say things, whether it's right or wrong, it's really fascinating.

10.20 There was some excellent homework, really excellent, normally I don't give out commendations for just one piece of work, normally I give out one commendation for three A's or A minuses but some pieces of work were really excellent. There were only a few Bs. No one got lower than a B. But there were some mistakes I didn't expect. Why shouldn't there have been some of those mistakes from a top set? Not because you're all brilliant and know it all? Why?

P because they were for

Because they were on the board

teacher asks how do you say I am

pupil replies j'ai

teacher at a common mistake-I am told, I am small

pupil says je suis

teacher what's the il and elle bit?

Pupil il est /elle est

Teacher writers on the board

Teacher had you say I have

Pupil je

Teacher – j'ai not je, j'ai

All the time the teacher's writing these up onto the board one: the first person and third person:

Je suis	il/elle est
J'ai	il / elle a
Je m'appelle	il / elle s'appelle
Je porte	il / elle porte

You need that for your age in French, in French you have to say I have your age.

how do you say my name is?

Je m'appelle, double p double l

T = he / she?

P = il / elle

Had use they are aware?

Pupils put hands up

Teacher we have these pupils before anyone else?

What's the il elle bit? What the ending? Those are really really important.

The other thing is adding an s to eyes. What you have to do that? Are eyes masculine or feminine? the teacher offers an explicit English description of why cheveux is plural

When do you put a letter e on the end?

Another teacher rise in the room and speaks German. Teacher A replies in French.

Teacher says some of Miss Gill's class were amongst the best of the homework candidates, even though they don't know as much. Why could this be? Pupil suggest they knew that that topic. Teacher no, they didn't know that topic, they've never seen before, because they were trying that's why

It shows we have to be careful so we get your writing as perfect as you speaking.

OK: où habites-tu? P north Baddersely

Teacher asks comment ca s'écrit? This question-and-answer repeat itself for about five turns. Pupils saying where they live teacher asking how the places spelt

T – dans Avantage à la page 5. There is no talking, the pupils get out their books quietly.

T – il y a 6 photos, sous les photos il y a 6 phrases qui expliquent. T reads out the places in French. Alors 2 minutes avec un partenaire – quelles phrases vont avec quelle photo? The pupils have started moving towards each other whilst the teacher was explaining this and they have started to talk. All more or less on task, one couple look a little lost. Some talking in English for example that one goes with that one, the third one

10.36 Class back together now. The teacher asks numéro un, c'est où ça? pupil tried to read out the long word the teacher says I know that a long word that don't get confused. The teacher recasts. Pupils do not repeat. Pronunciation is not corrected.

Teacher il y a beaucoup d'industrie. The teacher goes through the exercise with the pupils. T – où sont les Antilles. P is it in the Caribbean? Yes – it's in the West Indies

Teacher gives instructions in target language. Vous allez écouter la cassette vous écrivez 2 choses, le nom et l'endroit. Teacher get people to translate the instruction to the pupils start getting their stuff out. Teacher disciplined class that they didn't listen to instructions when they were being translated. Next week I won't do this but I'm not sure now that everybody understands the target language instructions yet. The teacher explained in English- we don't want sentences, just keywords.

Tape: je m'appelle Amelie. J'habite a ...La Guadeloupe eest un ile près de . plus some more French - not relevant to task they must do.

A few pupils say what? What?

T on va faire ca comme exemple.

Teacher writes the answers on the board. Okay you've got the idea, are getting both bits. Each extract is listened to twice. Six people on the cassette. The pupils must hear the keywords ie the name, town, place. By listening in reading the text on page four-actually they can spot the name and then read to find the rest of the information. Some are working together, slight muttering, but mainly each working individually. The teachers says numero 3 etc before each extract to help learners keep where they are.

T – OK comment elle s'appelle? The pupil just gives the name not il s'appelle. The teacher makes clear all I want it if you've got the main area somewhere in that area. The teacher read correct answer on the board. Continues getting the pupils to give answers. They give the name and place only. [why not put the names mixed up on the board, not in the correct order , so they have to give complete sentences with the subject and the verb, and the teacher could ask the pupils to give her the answers in any order, so the pupils have to specify who they are talking about!].

T – il y a 12 reponses – qui a douze sur douze? Almost all pupils put their hand up nobody says that they got 11 or ten. Moins? Nobody? Be pleased, put your hand up show me

Teacher moves onto the next activity page 5. Il faut decider qui. Vous avez regardé les photos, vous avez écouté la cassette. Maitenant c'est facile. It easy. They must write down where the people live page 5 exercise c.

There are some puzzled looks. All the people settle down. Teacher says you have all the information you need. Very calm atmosphere-almost silent teacher goes round the room checking pupils are OK. T – tu as fini? P – non T – pas encore. Teacher comments to the people next Amir alone today, you've lost your bodies. Encourages the pupil-good. Two to get round check-in work.

Teacher OK – qui habite a , oh, il , it helps you there, it's got to be a boy, pupils give just the place, their sentences. The book gives 1) il habite à Guadeloupe – and pupils must give the names by looking on the page opposite.

10.55 The teacher gives all instruction in the target language. Vous allez donner un detail. Le partenaire va deviner. Older pupils are attentive, quiet listening. Okay? How's it going? (the implication is that there is some shared understanding that the teacher is going to be giving instructions in the TL more and more and she is encouraging them because they are trying to listen. One pupil offers what you think the instructions might have meant in English, almost 100 percent correct. Teachers which side of the board is going to help? Pupils save the right side. Teacher rubs off the 'je' saide of the verb list, leaving 3rd person singular forms on there.

One pupil leaves the cheques is OK with teacher first. Teacher sorts out so that no pupils are working alone, pupils working groups of three rather than work alone.

I wander round the room listening to Paris, I hear about six pairs. Many pupils appear to be using the first person, as this is reading straight out of the textbook. They're picking extracts from the textbook and the pupil has to guess which person their partners talking about. 5 using je, 7 using third person plural (they do not **have to** use the third person for the activity). A lot of English being used to negotiator whose turn it is , clarifying answers etc, there is confusion between il a and il est. T – n'oublie pas, don't forget....[?]

One pupil starts j'ai, oh no j'ai, looks at board, il a...."

Teacher asks how many used il habite? And well done if you used il s'appelle rather than just saying the name. Teacher je vais expliquer les devoirs. There are lots of jobs (English). Teacher read down the list of vocabulary in French il est, elle est [jobs in Avantage]. Teacher asks what's the point? Pupil offers it's the difference between male and female. Teacher read down more jobs, this time with both versions, masculine and feminine. Teacher asks what it is only one ending? Pupil offers because its originally come from a male job? Teacher yes, I would think most of these were originally done by men, but say secretaire? Why is there only one form? Is it that originally they were all women. P suggests because there is an e on the end. Teacher what does an e on the end normally tell you? P – that it's feminine. Yes an e normally means that it's feminine. [all in English] so we can't add another. Okay let's leave that. I explain in French. Il y a 26 exemples, il faut noter le vocabulaire, en français en anglais, il faut trouver al forme, s'il y un forme feminine, et puis après en vert, il y a les endroits où ils travaillent.

Trouvez un endroit pour ...exemple le prof travaille dans un college, le secretaire...who understands? Two hands go up. (P1 – same boy that tried to explain activity before??) [he almost explained everything]. P explains in English. Pupils now start to get their homework diaries out and start to write down their homework. [this is real communication – they NEED to know their homework, target language opportunity?] another pupil tries to explain. Teacher says they can just think through where each person works – they don't need to write the sentence, just make sure they have copied vocab down and can make a sentence for each place (ie no verbs in homework). Due in on Friday 21st. Make sure you've noted vocab, even if you don't write other bit.

11.10 OK c'est très bien. Rangez les affaires. Les chaises sous les tables (said whilst pupils pack up and chat). Pupils stand and wait behind chairs – teacher says to a few at front they can go and rest follow.

Discussion afterwards with teacher

All have come from teacher X – most from the class I followed in action research project and others from the parallel set. Just 4 from another teacher – they have been placed between buddies from teacher X's group - each one with two of teacher X's on either side.

T expresses concern that looking through books they have not done much free writing, they are mainly grammar, paradigm, translation, picking out chunks, copying, and vocab written up – not much marking done of pupils' work. Says she intends to allow them to do a lot of writing and correcting their work.

Appendix 12 Example of a PI unit

or Tuesday 22nd - session 1

her - 'EI L+R activities' (audio recorded)

M - 'processing instruction' (video!)

Tape on Monday!

je will call the two groups

Blue (L + R activities) e playing

Green (processing instruction) nous ~~commence~~ - we listen to (or we are listening to)

Nous ne restONS pas = we don't stay (or we aren't staying)

What goes wrong?

This can be difficult for learners of French. Very often they do not learn to use this ending.

They say and write things like *nous jouer*.

Why?

It is easy to ignore this 'ons' ending, because once you have read or heard 'nous' you do not pay attention to the end of the verb.

It may be even harder to notice *ons* when the verb is negative:

Nous ne jouONS pas au squash.

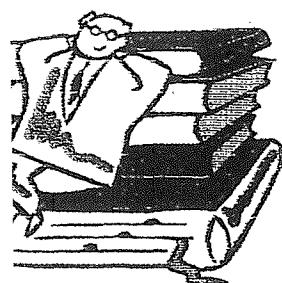
You must learn to pay attention to the ending of the verb 'ons', as well as the word 'nous'.

These activities will help you to do that.

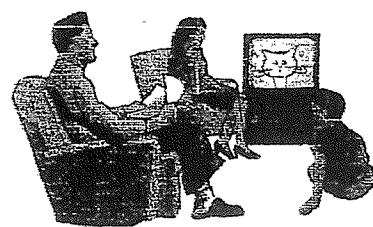
Some Hints!

CHT - show this and read it out as they read it. It should be very brief, exactly as written here - simply stick to the CHT!

verbs in French, you must have 'ons' on the end of



Le Soir d'une Professeur



Listen to this teacher talking about what normally happens in her evenings.

Put a tick in the 'je / I column' if she is talking just about herself,

or

a tick in the 'nous / we column' if she is talking about herself and other people.

You will hear only part of each sentence. You will not hear the word for 'I' or 'we'.

Remember: listen for the 'ons' ending to tell you if she is talking about 'we'.

<i>Sentence</i>	<i>Je / I</i>	<i>Nous / we</i>
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		

Le Soir d'une Professeur.

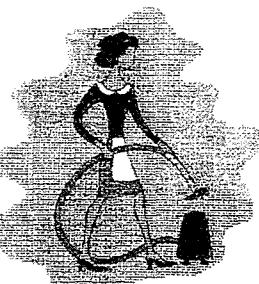
Transcript

- 1 Joue au tennis
- 2 Écoute la radio
- 14 Ne regardons pas la télé
- 3 Préparons le dîner
- 4 Travaille dans le bureau
- 5 Parlons au téléphone
- 6 Rentrons à 6 heures
- 7 Mangeons
- 11 Ne lavons pas la vaisselle
- 12 Ne corrige pas les cahiers
- 13 Ne range pas la maison
- 8 Fumons des cigarettes
- 9 Donne à manger au chien
- 10 Promenons le chien
- 15 Allons au lit à 11 heures

5 negatives

9 nous

6 je



La Vie Moderne:

Partageons le ménage à la maison... ?



70 women have written about the jobs that they do by themselves ("I...") in the house and those that are shared with their partners ("we...").

Qui fait le plus seul?

Coche les tâches qu' elle fait **seule**.

Qui fait le plus seul?

Coche les tâches qu' elle fait **seule**.

trice

/ nous...

range la chambre

lave la vaisselle

allons au supermarché

préparons le dîner

promène le chien

change les draps sur les lits

aide les parents au supermarché

travaillons dans le jardin

passons l'aspirateur

recycle le papier et les bouteilles

trice

/ nous...

écrivons la liste pour les courses

sortons la poubelle

aisons les courses

répare le dîner

lavons la vaisselle

éparons la voiture

promène le chien

range la chambre

passee les chemises

nettoyons les fenêtres

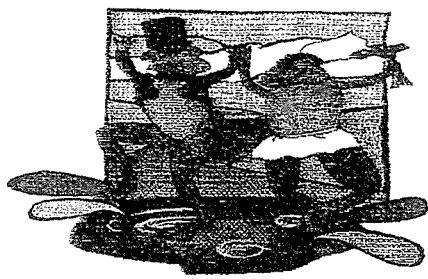
_____ fait le plus toute seule!

_____ et son partenaire partagent le ménage le plus.



Comment passes-tu le weekend?

Seul ou avec des autres?



Listen to 2 people talking about what they normally do at the weekend.

- Do you think they do these activities alone (seul) or with others (avec des autres)?

(remember: listen for the 'ons' ending!)

Decide which speaker you think might prefer being by themselves.

- You must also indicate whether **you** do this activity often, sometimes or never.

Underline the appropriate word in the 'moi' column.

Decide which speaker you would get on with best.

<i>Speaker 1</i>	<i>Probablement seul</i>	<i>Probablement avec des autres</i>	<i>Moi</i>
1			souvent / quelquefois / jamais
2			souvent / quelquefois / jamais
3			souvent / quelquefois / jamais
4			souvent / quelquefois / jamais
5			souvent / quelquefois / jamais
6			souvent / quelquefois / jamais
7			souvent / quelquefois / jamais

Speaker 2

1			souvent / quelquefois / jamais
2			souvent / quelquefois / jamais
3			souvent / quelquefois / jamais
4			souvent / quelquefois / jamais
5			souvent / quelquefois / jamais
6			souvent / quelquefois / jamais
7			souvent / quelquefois / jamais

Je trouve que numéro 1 / 2 * préfère être seul.

Who do you think you would get on with best? _____

* delete as appropriate

Transcript:

Speaker 1:

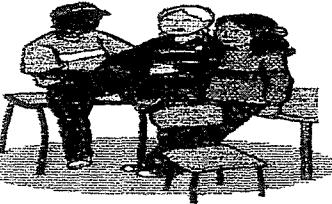
1. mangeons au restaurant une fois à la semaine
2. regardons la télé tous les soirs
3. écoute la radio dans la cuisine
4. parlons aux amis beaucoup
5. reste à la maison de temps en temps
6. nageons dans le centre sportif une fois à la semaine
7. travaillons quelquefois dans le jardin

Speaker 2:

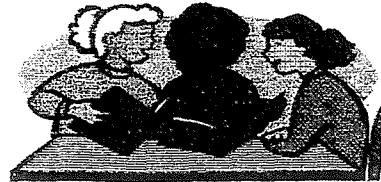
1. pratiquons un peu de sport
2. aime lire
3. prépare le dîner
4. écoute la radio
5. visitons le parc de temps en temps
6. joue du piano
7. promenons le chien

Speaker 1: 5 nous, 2 je

Speaker 2: 3 nous, 4 je



Des Profils



Comme Nous??

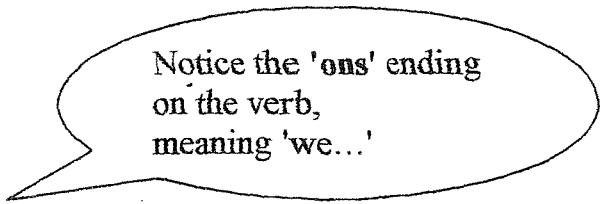
This "Opinions Profile" was written by a group of school-aged friends in France looking for some friends in England to write to and visit.

Coche les phrases si toi et tes amis avez les mêmes opinions.

Tick the sentences if you & your friends have the same opinions.

Nous...

- 1) détestons les maths
- 2) trouvons le football super
- 3) pensons que le roller-blading est cool
- 4) allons au cinéma toutes les semaines
- 5) adorons manger chez MacDonalds
- 6) jouons sur l'ordinateur
- 7) parlons au téléphone beaucoup



Notice the 'ons' ending on the verb, meaning 'we...'

Nous ne / n'...

- 8) faisons pas les devoirs
- 9) écoutons pas beaucoup aux professeurs!
- 10) aimons pas l'école

Discute avec le groupe si vous avez les mêmes opinions



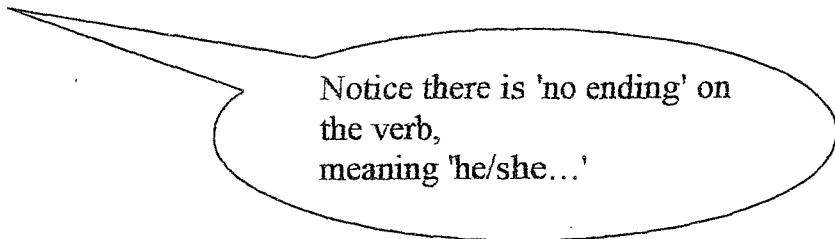
Another profile is from a young person
who is looking for a date!



Coche les phrases si tu as les mêmes opinions

Il / Elle...

1. aime les devoirs
2. joue beaucoup au tennis
3. mange souvent aux restaurants
4. protège l'environnement
5. recycle toujours le papier
6. préfère le chocolat aux carottes
7. adore le cyclisme



Notice there is 'no ending' on the verb, meaning 'he/she...'

Il / Elle ne...

8. travaille pas les weekends
9. danse pas
10. regarde pas les actualités (news) à la télévision

Est-ce que tu as les mêmes opinions que ton partenaire?

Appendix 13 Example of an EI unit

Tuesday 22nd - session 1

her - 'EI L+R activities' (audio-recorded)

M - 'processing instruction' (videoed!)

Tape on Monday!

we will call the two groups
Blue ('EI' L+R activities) e playing

Green (processing instruction)
nous - we listen to (or we are listening to)

Nous ne restONS pas = we don't stay (or we aren't staying)

What goes wrong?

This can be difficult for learners of French. Very often they do not learn to use this ending.

They say and write things like *nous jouer*.

Why?

It is easy to ignore this 'ons' ending, because once you have read or heard 'nous' you do not pay attention to the end of the verb.

It may be even harder to notice *ons* when the verb is negative:

Nous ne jouONS pas au squash.

You must learn to pay attention to the ending of the verb 'ons', as well as the word 'nous'.

These activities will help you to do that.

Some Hints!

OHT - show this and read it out as they read it. It should be very brief, exactly as written here - simply stick to the OHT!

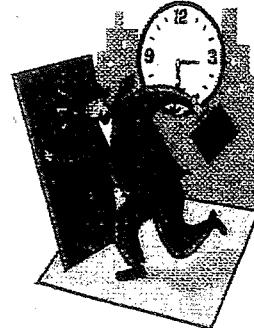
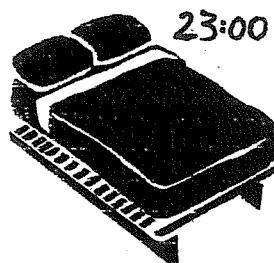
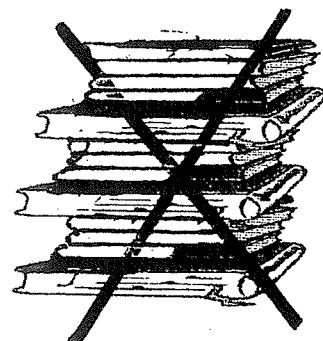
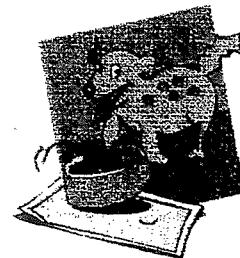
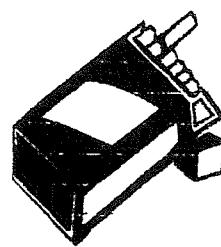
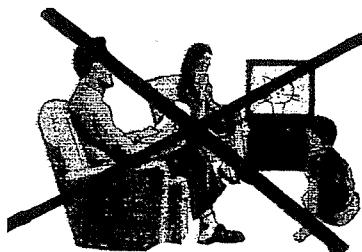
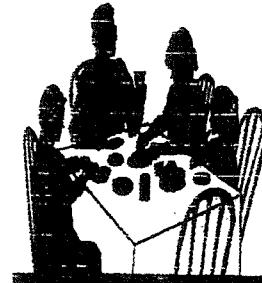
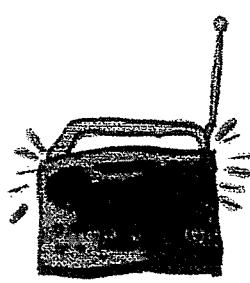
rbs in French, you must have 'ons' on the end of

The tape plays this activity twice! No rewinding necessary. After the second time, show them the OHT with the correct order written on. They simply connect their **Le Soir d'une Professeur** own. Ask for marks:

"qui a 10? 11? 12? 13? etc"

Écoute la professeur, qui parle d'un soir typique.

Écris l'ordre des activités. (Écris 1 à côté du premier dessin etc.) just to check the vocab. underlined in the transcript.
Please do not stop it for other reasons.



Transcript

- 1 À 5 heures je joue au tennis ou squash, normalement 3 fois à la semaine.
- 2 J'écoute la radio dans la voiture quand je rentre chez nous.
- 3 Nous rentrons tous les deux à la maison à 6 heures quinze.
- 4 Je donne à manger au chien qui a toujours faim.
- 5 Et après nous promenons le chien
- 6 Puis nous préparons le dîner tout de suite parce que nous avons faim.
- 7 Nous mangeons dans la salle à manger, toute la famille ensemble.
- 8 Je ne lave pas la vaisselle
- 9 Nous ne regardons pas beaucoup la télé ~~■■■~~
- 10 ~~Mais~~ Après le dîner, je travaille dans le bureau sur l'ordinateur.
- 11 souvent, je ne corrige pas les cahiers – je suis trop fatigué.
- 12 Nous ne rangeons pas la maison non plus.
- 13 Nous parlons au téléphone, aux amis ou à la famille.
- 14 Nous fumons des cigarettes
- 15 Normalement nous allons au lit à 11 heures

5 negatives

9 nous

6 je

Pupils work on this activity alone.
Then you ask 'numéro 1, vrai ou faux?'
Volunteers say either 'vrai ou faux.'

La Vie Moderne:

Partageons le ménage à la maison...?

Read what 2 women have written about the jobs that they do by themselves in the house and those that are shared with their partners. Decide whether the statements below are true or false.

Marie

Moi, je range la chambre et aussi je lave la vaisselle. Et c'est moi qui change les draps sur les lits.

Mais moi et mon mari, Marc, nous allons au supermarché et nous préparons le dîner ensemble. Tous les soirs je promène le chien toute seule. Le weekend, nous travaillons souvent dans le jardin et nous passons l'aspirateur. Moi, je veux protéger l'environnement alors je recycle le papier et les bouteilles.

Mes parents habitent près d'ici alors je les aide faire les courses au supermarché.

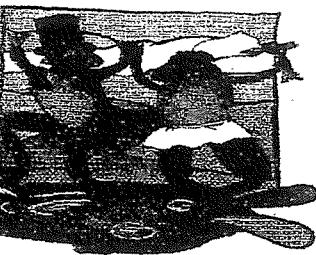
Hélène

Moi et mon mari Jacques, nous écrivons la liste pour les courses, et nous sortons la poubelle tous les deux. En plus, nous faisons les courses ensemble. Mais, normalement, je prépare le dîner. La plupart du temps nous lavons la vaisselle. Même nous réparons la voiture ensemble! Pourtant, je promène le chien, je range la chambre et je repasse les chemises. C'est vrai que nous nettoyons les fenêtres ensemble.

	Vrai	Faux	Pas dans le texte
1. Marie et Marc vont au supermarché	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Marie ne travaille pas dans le jardin	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
3. Marie aide ses parents à faire les courses	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Marc passe l'aspirateur	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Marc ne nettoie pas les fenêtres	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
6. Marie n'aime pas les voitures	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
7. Jacques prépare le dîner	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
8. Hélène promène le chien	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Hélène et Marc repassent les chemises	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
10. Hélène ne range pas la chambre	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
11. Hélène donne à manger au chien	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
12. Jacques ne lave jamais la vaisselle	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Read instructions ~~1~~ to the pupils

Each number is played twice on the tape.



Comment passes-tu le weekend?

Seul ou avec des autres?

Listen to 2 people talking about what they normally do at the weekend.

- Do you think they do these activities alone (seul) or with others (avec des autres)?

There may not be a right or wrong answer

Decide which speaker you think might prefer being by themselves.

- You must also indicate whether **you** do this activity often, sometimes or never.

Underline the appropriate word in the 'moi' column.

Decide which speaker you would get on with best

Say:
It doesn't matter
whether you do it
alone or with other
people - just whether
you do the activity

Speaker 1	Probablement seul	Probablement avec des autres	Moi
1			souvent / quelquefois / jamais
2			souvent / quelquefois / jamais
3			souvent / quelquefois / jamais
4			souvent / quelquefois / jamais
5			souvent / quelquefois / jamais
6			souvent / quelquefois / jamais
7			souvent / quelquefois / jamais

Speaker 2

1			souvent / quelquefois / jamais
2			souvent / quelquefois / jamais
3			souvent / quelquefois / jamais
4			souvent / quelquefois / jamais
5			souvent / quelquefois / jamais
6			souvent / quelquefois / jamais
7			souvent / quelquefois / jamais

Je trouve que numéro 1 / 2 * préfère être seul.

Who do you think you would get on with best? _____

* delete as appropriate

Transcript:

Speaker 1:

1. Nous mangeons au restaurant une fois à la semaine avec des amis
2. Moi et ma famille, nous regardons la télé tous les soirs
3. J'écoute la radio dans la cuisine quand je prépare le dîner
4. Nous parlons aux amis beaucoup
5. Je reste à la maison de temps en temps
6. Nous nageons avec les enfants dans le centre sportif une fois à la semaine
7. Moi et mon mari, nous travaillons quelquefois dans le jardin

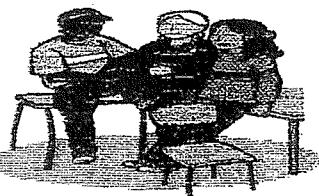
Speaker 2:

1. Nous pratiquons un peu de sport avec les amis
2. J'aime lire dans ma chambre
3. Je prépare mon dîner
4. J'écoute mes disques compacts
5. Nous visitons le parc de temps en temps avec des amis
6. Je joue du piano
7. Nous promenons le chien

Speaker 1: 5 nous, 2 je

Speaker 2: 3 nous, 4 je

This is a quick activity - the "discussion bit" in italics is just in case the other group hasn't come back yet!



Des Profils



Comme Nous??

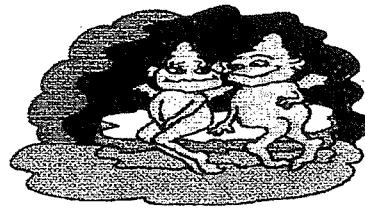
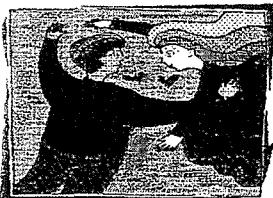
This "Opinions Profile" was written by a group of school-aged friends in France looking for some friends in England to write to and visit.

Coche les phrases si toi et tes amis avez les mêmes opinions.
Tick the sentences if you & your friends have the same opinions.

- 1) Nous n'aimons pas l'école
- 2) Nous détestons les maths
- 3) Nous n'écoutes pas beaucoup aux professeurs!
- 4) Nous trouvons le football super
- 5) Nous pensons que le roller-blading est cool
- 6) Nous ne faisons pas les devoirs
- 7) Nous allons au cinéma tous les semaines
- 8) Nous adorons manger chez MacDonalds
- 9) Nous jouons sur l'ordinateur
- 10) Nous espérons aller à l'université

Discute avec le groupe si vous avez les mêmes opinions

Another profile is from a young person
who is looking for a date!



Coche les phrases si tu as les mêmes opinions.

1. Il / Elle ne travaille pas les weekends
2. Il / Elle aime sortir avec les amis
3. Il / Elle joue beaucoup au tennis
4. Il / Elle mange souvent aux restaurants
5. Il / Elle pense qu'il faut protéger l'environnement
6. Il / Elle recycle toujours le papier
7. Il / Elle préfère le chocolat aux carottes
8. Il / Elle adore faire du cyclisme
9. Il / Elle ne danse pas
10. Il / Elle ne regarde pas les actualités (news) à la télévision

Est-ce que tu as les mêmes opinions que ton partenaire / le reste du groupe?

Appendix 14a: Record of intervention protocol

Class A	Lesson	Example of linguistic feature	Date	PI teacher	EI teacher	Note		
1	1	je joue & nous jouons	Mon 21 Jan	EM	video	A	audio	EI activities completed but not handed in
2	2	il a & il est after intervention, whole class activities for about 12 mins	Wed 23 Jan	A	video	EM	audio	12 mins whole class non-intervention activities
3	3	il (+je) joue & ils jouent (reading only) il est, a, fait, va & ils sont, ont, font, vont	Mon 28 Jan	EM	video	A	audio	last EI activity (out of 6) not done
4	4	Teacher A takes whole class	Fri 1 Feb	whole class non-intervention activities				
5	5	Teacher A absent - whole class cover work	Mon 4 Feb	whole class non-intervention activities				
6	6	il joue & tu joues tu joues & vous jouez after intervention, whole class activities for about 12 mins	Wed 6 Feb	A	video	EM	audio	12 mins whole class non-intervention activities
7	7	joue/jouons & joué; je joue & j'ai joué	Mon 18 Feb	EM	audio	A	video	
8	8	tu joues & tu as joué il joue & il a joué Teacher A absent - cover work	Fri 22 Feb	EM	audio	EM	audio	25 mins non-intervention activities
9	9	nous jouons & nous avons joué vous jouez & vous avez joué	Mon 25 Feb	EM	audio	RM	video	
10	10	tu as & il a joué (reading only) Teacher A absent - cover work	Wed 27 Feb	EM	audio	EM	video	25 mins non-intervention activities
11	11	il a mangé & il est allé il a mangé/est allé & ils ont mangé/sont allés	Mon 4 March	EM	video	A	audio	

Appendix 14b: Record of Intervention protocol

Class B Lesson	Language forms	Date	PI teacher	EI teacher	Note
1	je joue & nous jouons	Tues 22 Jan	EM video	B audio	last half of last EI activity (out of 4) not done by most
2	il a & il est	Tues 29 Jan	B video	EM audio	
3	il (+je) joue & ils jouent (reading only) il est, a, fait, va & ils sont, ont, font, vont	Fri 1 Feb	EM video	B audio	last EI activity (out of 6) not done v thouroughly by all
4	il joue & tu joues tu joues & vous jouez	Tues 5 Feb	B video	EM audio	
5	joue/jouons & joué; je joue & j'ai joué	Tues 19 Feb	EM audio	B video	
6	tu joues & tu as joué il joue & il a joué tu as & il a joué (reading only)	Thurs 21 Fe	B video	EM audio	
7	Teacher B absent - cover work ('au camping')	Fri 22 Feb	whole class non-intervention activities		
8	nous jouons & nous avons joué Teacher B absent - cover work ('au camping')	Tues 26 Feb	EM video	EM audio	
			20 mins non-intervention activities		
9	vous jouez & vous avez joué	Tues 5 Mar	B video	EM audio	
10	il a mangé & il est allé il a mangé/est allé & ils ont mangé/sont allés	Thurs 7 Mar	EM video	B audio	

Appendix 15: Analyses to assess the randomised matched pairs sampling procedure

		Sum of Squares	df	Mean Square	F	Sig. ¹
LIST	Between Groups	223.616	2	111.808	.560	.574
	Within Groups	16383.153	82	199.795		
	Total	16606.769	84			
READ	Between Groups	3235.886	2	1617.943	4.790	.011
	Within Groups	27699.062	82	337.793		
	Total	30934.948	84			
WRIT	Between Groups	5698.081	2	2849.041	7.974	.001
	Within Groups	29298.922	82	357.304		
	Total	34997.003	84			
SPK	Between Groups	551.177	2	275.589	1.481	.238
	Within Groups	8559.555	46	186.077		
	Total	9110.732	48			

One way ANOVAs to compare the pre test scores for PI & EI merged groups and class C

	(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	95% Confidence Interval	
					Lower	Upper
Read	PI	EI	-3.1928	4.99655	.799	-15.2452 8.8595
	Class C	EI	11.0030	4.96440	.078	-9637 22.9698
	Class C	PI	14.1959	4.72260	.011	2.8252 25.5666
Writing	EI	PI	.7238	4.16666	.984	-9.3238 10.7715
	Class C	EI	16.7405	5.16372	.006	4.2540 29.2269
	Class C	PI	17.4643	5.34126	.005	4.5783 30.3503

Post hoc tests² to compare the scores from the merged EI & PI groups and class C at pre test

Test measure	class	t^3	df	Sig. (2-tailed)	Mean Difference of the Difference	Std. Error of the Difference	95% Confidence Interval	
							Lower	Upper
List	A	-.415	26	.681	-2.6793	6.45111	-15.93972	10.58115
	B	.303	25	.764	1.1790	3.89209	-6.83690	9.19492
Read	A	.581	26	.566	3.8964	6.70887	-9.89384	17.68670
	B	.503	25	.620	1.5310	3.04666	-4.74373	7.80571
Writing	A	-.331	27	.743	-2.2002	6.65405	-15.85322	11.45274
	B	1.235	24	.229	3.3410	2.70434	-2.24054	8.92244
Speaking	A	-.034	14	.973	-.2733	8.07304	-17.58830	17.04161
	B	-.054	15	.958	-.1699	3.13626	-6.85463	6.51493

Independent samples t tests to compare the PI and EI groups at pre test in all measures

¹ For the listening, reading and speaking test measures, Levene's test of homogeneity of variances suggested that the variances of the groups' scores at pre test could be considered normal (appendix 32). For the writing test scores, it was found that the results of the ANOVA should be interpreted with caution. However, non-parametric tests were carried out and, as they supported the findings presented below, they are not shown here. Non parametric tests were also carried out for the speaking tests as the n was quite small, though again they are not presented here as they supported the parametric tests. See section 3.8 for presentation and discussion of the statistical procedures used.

² The Games Howell post hoc test was used, as samples had unequal variances and sizes.

³ Equal variances could be assumed according to Levene's test for all variables.

Appendix 16 Sequence of Linguistic Features

Unit	Example	Feature
<i>Present tense - Number / person inflections</i>		
1	je joue & nous jouons (+ in neg.)	number
2	il a & il est (+ in neg.)	semantic, avoir + être
3	il (+je) joue & ils jouent (+ in neg.) il est, a, fait, va & ils sont, ont, font, vont (+ in neg.)	number number
4	tu joues & vous jouez (+ in neg. + interrog.)	number
<i>Inflections for present & perfect tenses</i>		
5	joue/jouons & joué; je joue & j'ai joué (+ in neg.)	past participle aux - (inc. 'fait' type)
6	tu joues & tu as joué il joue & il a joué (+ in neg.) tu as & il a	aux - tense aux - person
7	nous jouons & nous avons joué (+ in neg.)	aux - tense
8	vous jouez & vous avez joué (+ interrog.)	aux - tense
<i>Perfect tense - Number / person inflections</i>		
9	il a mangé & il est allé (+ in neg.)	semantic/arbitrary (avoir & être)
10	il a mangé & ils ont mangé (+ in neg.) il est allé & il sont allés (+ in neg.)	number

Appendix 17 Ensuring the two versions of the tests obtained similar results

Although only the pretest results are shown, the post and dp tests obtained similar results.

	Test version	N	Mean	Std. Deviation	Std. Error Mean
List	1	58	50.6460	14.68547	1.92830
	2	27	40.8948	9.92695	1.91044
Read	1	42	45.4545	21.87348	3.37515
	2	43	40.8028	16.07925	2.45206
Writ	1	41	24.7854	23.95579	3.74127
	2	43	19.2878	16.36294	2.49533
Spk	1	22 ¹	21.3259	15.34092	3.27070
	2	21	17.3643	7.51078	1.63899

Descriptive statistics

	Test version	Kolmogorov-Smirnov statistic	df	Sig.	Shapiro-Wilk statistic	df	Sig.
Listening	1	.099	58	.200	.980	58	.150
	2	.124	27	.200	.987	27	.167
Reading	1	.132	41	.068	.931	41	.015
	2	.140	43	.034	.950	43	.058
Writing	1	.218	41	.000	.813	41	.000
	2	.164	43	.005	.876	43	.000
Speaking	1	.206	22	.016	.812	22	.001
	2	.122	21	.200	.981	21	.942

Assessing the normality of the distributions

	F	Sig.	t^2	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Listening	4.656	.034	3.592	71.915	.001	9.7512	2.71443	4.34000	15.16244
Reading	3.001	.087	1.119	83	.266	4.6517	4.15705	-3.61646	12.91993
Writing	3.087	.083	1.233	82	.221	5.4976	4.45782	-3.37041	14.36566
Speaking	4.396	.042	1.083	30.830	.287	3.9616	3.65838	-3.50142	11.42454

Independent samples t tests for equality of means

	READPRE	WRITPRE	SPKPRE
Mann-Whitney U	836.000	789.500	219.500
Wilcoxon W	1782.000	1735.500	450.500
Z	-.590	-.823	-.280
Asymp. Sig. (2-tailed)	.555	.410	.780

Non parametric statistics

¹ For the analysis of the speaking test versions, 6 learners (3 from class A, 2 from class B and 1 from class C) took different versions of the sentence and narrative tests (e.g. version 1 of the sentence level but version 2 of the narrative tasks) and so these were excluded from this analysis. This left 43 learners in this analysis.

² Equal variances could be assumed for the reading and writing tests but not for the listening and speaking tests. SPSS provided the suitable test statistic taking this into account.

Appendix 18 Breakdowns of Linguistic Features in Interpretation Tests

Interpretation of written inflections

Task no.	Description of test	Tense	Feature details <i>all declarative</i>	No. items (excl. distractors)
1	Choosing subject for inflected verbs.	present	Er verbs	12
2	Choosing subject for inflected verbs.	perfect	Er verbs, avoir, all persons	12
2	Choosing subject for inflected verbs.	perfect	Er verbs, être, 3 rd pers, sing & plural	4
4	overriding event probabilities, matching infl to subject.	present	Er verbs, all except 1 st pers. singular	5
5	overriding event probabilities, matching infl to subject.	perfect	Er verbs, avoir, all except 1 st pers. sing.	5
6	matching verb to picture	present vs perfect	Er, avoir	11
6	matching verb to picture	present vs perfect	'Non er' (aux only possible cue écrit, fait).	3
6	matching verb to picture	present vs perfect	Er, être, 3 rd person sing & plural only	1 (2 in version 2)
6	matching verb to picture	present vs perfect	'Non er', être, 3 rd person sing & plural	2 (1 in version 2)
				Total = 55

Interpretation of aural verb inflections

Task no.	Description of test ¹	Tense	Feature details <i>declarative unless specified</i>	No. items (excl. distractors)
1	Choosing subj. for infl. verbs	present	Er, null, ons, ez	3
1	Choosing subj. for infl. verbs	present	Er, negative, null, ons, ez	3
2	Choosing subj. for infl. verbs	perfect	Er, ai, a, avons, ont	4
2	Choosing subj. for infl. verbs	perfect	Er, negative, ai, a, avons, ont	4
2	Choosing subj. for infl. verbs	perfect	Er, sont	1
2	Choosing subj. for infl. verbs	perfect	Er, negative, sont	1
3	Choosing sing. or pl.	present	est, a, va, fait versus sont, ont, vont, font	8
4	Choosing pres or past	present & perfect	Er verbs, avoir, all persons, aux & pp cues	12
4	Choosing pres or past	present & perfect	Er verbs, avoir, negative, 1 st & 3 rd pers sing, aux & pp cues	4
4	Choosing pres or past	present & perfect	Er verbs être, 3 rd person sing & plural only,	2
4	Choosing pres or past	present & perfect	Aux. <i>only</i> possible clue (e.g. fait, dit & vous mangez/mangé)	6
				Total = 48

¹ Learners could select 'not sure' for each item

Appendix 19 Breakdowns of Linguistic Features in Production Tests

Production of written verb inflections

Task	Description	Tense	Feature details	No. items (excl. distractors)
1	Gap fill	present	Er, all persons (3 rd pers sing - 2 nouns, 1 pron)	9
2	Gap fill	present	Est, a, va, fait, sont, ont, vont, font	8
3	Describe weekend / hobbies sentence/discourse	present	Any verbs - not 2 nd person	8 (max.)
4	Gap fill	perfect	Er, avoir, all except avez	8
4	Gap fill	perfect	Er, être, 3 rd person sing & plural only	3
6	Describe weekend / holiday sentence/discourse	perfect	Any verbs - not 2 nd person	12 (max.)
				Total approx = 48

Production of oral verb inflections

Task	Description	Tense	Feature details	No. items (excl. distractors)
1	sentence Fill in blank	present	Er, regular, all persons	8
2	sentence Fill in blank	present	Est, va, a, fait & sont, vont, ont, font	8
3	sentence Fill in blank	perfect	Er, declarative, avoir	8
3	sentence Fill in blank	perfect	Er, declarative, être, 3 rd pers, sing & pl only	4
4	discourse Le weekend, narration	present	Er, declarative, targets: 1 st & 3 rd sing & plural	8 (max)
5	discourse Le weekend, 'Q & A'	present	Any, 2 nd , sing & plural	6 (max)
5	discourse Le weekend, 'Q & A'	present	Any, 1st, sing & plural	4 (max)
6	discourse Noël - narration	perfect	Any (er & fait, 3 rd person sing & plural, 1 st person plural targeted in example)	8 (max)
7	discourse Les vacances, 'Q & A'	perfect	Any, 2 nd , sing & plural	6 (max)
7	discourse Les vacances 'Q & A'	perfect	Any, 1st, sing & plural	6 (max)
				Total approx = 66

Appendix 20 Descriptions of the test activities

All instructions were read out to learners at every administration of the test, as well as learners being asked to read them on their sheets.

a) *Reading - Six activities - Appendix 21a*

1-4) *Four 'banked (or matching) fill in the gap' sentence level activities.*

- 2 offered a bank of different lexical verbs with a range of inflections
 - 6 target present tense *er* verbs
 - 8 target auxiliaries in the perfect tense (all *avoir* and singular & plural 3rd person *être* auxiliaries)).
- 2 offered a range of pronouns/subject
 - 6 target items in the present
 - 8 target items in the perfect

There were two non-target feature 'distractors' in each activity (e.g. where learners had to select a noun). These were not included in the analysis.

The tests were timed to within about 15 seconds. Learners had 10 minutes 30 to complete the 2 present tense reading tasks and 10 minutes to complete the 2 perfect tense tasks.

Although alphabetical presentation of the possible words is often suggested {Alderson 2000 238 /id} it was not used here as this would mean that the order would have been different for the two versions of the 'same test'. By 'scattering' the items it was also considered likely that learners' attention would not be unnecessarily drawn to the fact that there was a list of subject pronouns in the bank. The position of the words for both versions of the test was identical.

5 & 6) *Two sentence level subject & verb matching activities*

- 5 in the present
- 5 in the perfect

Pupils were given 2 minutes 30 for the present tense task and 2 minutes 15 for the perfect tense task.

Learners were asked to combine the start and ends of 5 sentences, though 6 possible endings were given. This task was designed to mislead learners to process verbs on the

basis of their lexical semantics (rather than by matching the inflection to the subject). The verb inflections went against event probabilities e.g. learners should have combined *Posh Spice* with *mange les carottes* rather than with *danses, chantes et composes de la musique*.

b) Listening - 4 tasks - Appendix 21b

The tape was not stopped during the activities, so each test took the same time at each test administration. Each sentence was heard twice before moving on to the next one.

Recording was done with an unsophisticated lapel microphone and audio-cassette recorder. There was no prosodic emphasis given to the target inflections to enhance their acoustical salience (unlike in VanPatten 1990 p290). The speech rate would probably not be considered to be native.

1) Present tense inflections for person and number.

Learners heard 6 sentences in the present tense, containing regular er verbs with their subjects removed and were asked to circle either *il, nous, vous* or *not sure*. It was my voice (i.e. non-native speaker) recorded. Each type of inflection (null, 'ons', 'ez') was heard twice, once in a declarative sentence and once in a negative sentence.

2) Perfect tense inflections for person and number

Learners heard 10 sentences in the perfect tense without subjects, 8 of which had *avoir* auxiliaries and 2 had '*sont*'. Learners had to circle one of *je, il, nous, ils* or *not sure*. I recorded this activity. Each type of inflection was heard twice, once in a declarative sentence and once in a negative sentence.

3) 3rd person present tense common irregular forms

Learners heard 8 declarative sentences containing *a, est, va, fait, ont, sont, vont, or font* with no subjects. They had to tick either the singular or plural subject e.g. *le train, les trains* or *not sure*. A bilingual male teenager recorded the sentences under my direction.

4) Verb inflections for present versus perfect tense

Learners heard 24 sentences (containing a mix of 12 present and 12 perfect tense) - a break down of the linguistic features is given in appendix . Pupils had to tick either present, perfect or *not sure* (the rubrics explained these labels). On every administration, I also said "it's ok to tick the *not sure* column if you don't know, aren't 100% sure or think you can't tell". This activity was stopped after 12 items¹ and the pupils completed a short reading

¹ on every administration, except with about 7 learners who did the test individually.

activity before continuing with items 13-24. This was simply to give the learners a break from listening to 24 French sentences.

c) Writing - 5 activities - appendix 22a

Sentence level

- 1) This consisted of 9 present tense declarative sentences with temporal adverbs indicating present tense context but regular *er* verbs missing. An English infinitive was given at the end in brackets indicating which French verb could be used to fill in the blank. The learners 1 minute 30 seconds to revise a list of the nine target verbs in English with their French infinitive equivalents (and the irregular infinitives for the next activity, see below). The lists were different for pupils taking test 1 and test 2. Before the task began they were told "the words you have just revised will help you a bit but not entirely".
- 2) 8 3rd person present tense declarative sentences required 4 singular and 4 plural forms of common irregular verbs (*faire, aller, être, avoir*). English infinitives were given at the end of the sentence in brackets and the English and French infinitives had on the revision list prior to the previous activity.

The learners were given 5 minutes 30 seconds to complete this and the previous task.

- 3) 11 sentences with past tense temporal adverbs and a range of subjects (see appendix) required completion with appropriate forms of *er* verbs. The English infinitive was given at the end of each sentence in brackets. The gap indicating the missing language contained one continuous line, so as not to indicate that two words were needed. The learners had 1 minute 15 seconds to revise the words they would be required to use. The words "in the past" were stressed when reading out the instructions. Before the task they were told "the words you have just revised will help you a bit but not entirely". 3 verbs took *être* (2 singular, 1 plural). Learners had 4 minutes 30 to complete the task.

Discourse level

- 4) Present tense.

Learners were asked to write about typical weekend activities guiding them to make five obligatory references to 1st person singular and plural subjects and five to 3rd person singular and plural subjects. Half of the activity required the learners to use their imagination. Picture prompts were given to the learners to help them to think of things to say for which they may have known the vocabulary. The pupils had 8 minutes 30 to complete this task.

5) Perfect tense

Learners were asked to write 12 sentences about past holidays, of which 6 were prompted to be in the 1st person (3 singular, 3 plural) and 6 in the 3rd person (3 singular and 3 plural). Some of the prompts required the learners to use their imagination. The same picture prompts were given to the learners as in the previous task. The learners were given 8 minutes 30 to complete the task.

d) Speaking – 7 tasks - Appendix 22b

The learners had as long as they wanted to complete the activity, though no learner took longer than about 90 seconds to complete each activity. They did not stop recording once these three activities had started.

Sentence level tasks

1) Sentence level - present tense - regular

Learners were given 1 minute 40 seconds to familiarise themselves with the symbols that were going to be used in the task, their English meaning and the French infinitives. They then heard me recorded doing an example task, reading out eight similar sentences with the same lexical verbs they would have to use but with different inflections. The instructions were then read out to learners and I emphasised, for every test, the present tense context by saying "these sentences are about what a group of friends *normally* do after school when they go home and make themselves a snack. Or you can imagine that they are a list of instructions for how to make chocolate mousse". The learners then started the activity, each working at their own pace through the sentences. Pupils sat a couple of metres away from each other and every attempt was made to place pupils doing test 1 near pupils doing test 2, to try to reduce the temptation to repeat their peers' utterances.

2) Sentence level - irregular present tense

At every administration, the information in the speech bubble "the sentences only need either DO, GO, HAVE or BE (is/are)" was stressed and two examples were given, one from test 1 and one from test 2 e.g. "the first picture you would need to say 'he is rich' or for those with the other test 'she does the homework'".

3) Sentence level - perfect tense

The learners were given 1 minute to revise the symbols, English and French infinitives of the lexical items they would need. The instructions were read out to them and I stressed

the words "what people **did** during the **last** holidays". I then gave them an example in English e.g. "During the holidays my Dad worked in the garden".

Discourse level narration tasks

4) Present tense narration of pictures

This task was intended to elicit singular and plural 1st and 3rd person present tense verbs.

Learners first listened to an example of the activity using headphones. Four versions of the test were used: test 1 and test 2 and a male and female version, as the gender and activity of the person in each picture was different. For example, the female pupils doing test 1 were instructed to imagine that they were the girl in the first picture and the other people were their family or friends, as in the recorded example they heard.

5) Present tense prompted conversation

This activity was designed to elicit first and second person singular and plural verbs. It also may have decreased the artificiality of the learners' language and increased the cognitive load, as they were required to think of something to say, and possibly, think about genuine events in their own lives. The researcher was given possible answers they could give to the learners' questions and guidance about how to encourage the learners to speak (e.g. use pictures from the first task as a source of ideas). As the learners' rubrics told them what would be said to them, it is acknowledged that these tasks did not always demand genuine interaction.

6) Perfect tense narration

Again the learners heard an example of the narration of a series of photos of events over a Christmas holiday, with dates above each photo. They were told that the person describing what happened was a friend of theirs telling them what they did over Christmas. They were then asked to tell researcher what their friend had done (i.e. attempting to elicit third person singular and plural obligatory contexts).

7) Perfect tense prompted conversation

This was a similar format to the present tense structured conversation, designed to elicit contexts for first and second singular and plural verbs.

Appendix 21a The interpretation tests: reading tests

Reading

These sentences are about what people eat.

Choose a word to fill in the blanks.

1. Je/j' _____ les pommes

2. Dans ma famille, on est fou du _____, par exemple le cheddar ou le brie

3. Nous _____ les hamburgers

canottes

adorons

4. Ils _____ les gateaux

fromage

5. Mon lapin adore les _____ !

aliment

prépare

6. Vous _____ le pain

détestez

7. Tu _____ les tomates

mange

8. Elle _____ la pizza

préfère

recommander

accepter

These sentences are about what people do to protect the environment.

Choose a word to fill in the blanks.

1. Au supermarché je choisis les fruits _____

je/j'

organiques

2. _____ recycle les bouteilles

voulez

ils

3. _____ achetez le papier recyclé

4. _____ utilisent les aerosols non-CFC

je/j'

ils

5. Je vais en vélo, pas en _____

voulez

tu

nous

6. _____ protège l'environnement

7. _____ prenons une douche – pas un bain – pour économiser l'eau

adorer

8. _____ respecte la nature

important

nécessaire

Reading

Match the first part of each sentence (1-5) with an ending (A-F).

Le président	—	A	prépare la table
<i>mère:</i> "Tu..."	—	B	détestent l'école, en général
Les professeurs	—	C	corrigéons les cahiers tous les jours
Moi et mes amis, nous	—	D	parler beaucoup en français
<i>professeur:</i> "Très bien, vous..."	—	E	représentes le gouvernement
		F	fumez des cigarettes

These sentences are about what happened in the holidays last summer.
Choose a word to fill in the blanks. You can use the words more than once or not at all.

L'été dernier...

1. Ils _____ passé 10 jours en France est
2. Il y avait beaucoup de soleil et il faisait _____ tous les jours avons aurons
3. Tu _____ quitté l'hôtel à 11 heures détestais
4. Le premier jour, nous _____ nagé dans la mer sont
5. Vous _____ acheté trop de souvenirs chaud
6. Ma copine _____ rentrée en Angleterre le 25 juillet froid
7. Je/j' _____ commençé les vacances le 15 juillet ont
8. Mon ami se bronzait sur la _____ à
9. Oui, elle _____ visité *tous* les musées plage
10. Mes copains _____ arrivés en Espagne le 8 août avez suis

These sentences are about what people did last weekend in their free time.
Choose a word to fill in the blanks. You can use the words more than once or not at all.

Le weekend dernier...

1. _____ a planté des fleurs dans le jardin Marc
2. _____ avez corrigé les devoirs téléphone chien
3. On promenait le _____ dans le parc ils
4. _____ ai fait les courses au supermarché je // j'
5. Samedi soir, _____ ont rangé la maison travailler elle
6. _____ sont montés à cheval adorer
7. Normalement je parle aux amis au _____ vous détester
8. _____ as mangé du chocolat nous
9. Dimanche soir, _____ avons étudié pour l'université dimanche
10. _____ est resté à la maison pour faire les devoirs les enfants

Reading

Match the first part of each sentence (1-5) with an ending (A-F).

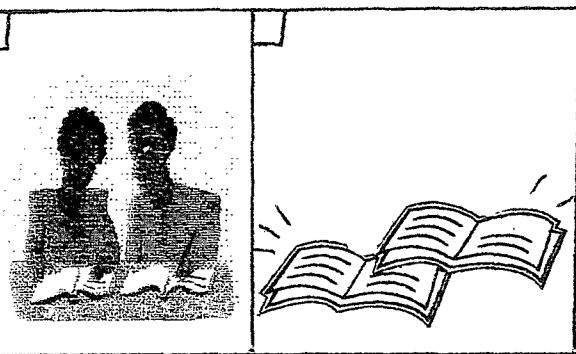
1.	Le directeur de l'école	_____	A	parler bien en français, super!
2.	<i>mère</i> : "Tu..."	_____	B	avons mangé chez Monica et Chandler
3.	Les acteurs de "Friends"	_____	C	as exclu (=expel) les élèves
4.	<i>professeur</i> : "Excellent, vous..."	_____	D	ont nagé à Romsey Rapids
5.	Le weekend dernier, nous	_____	E	a fait les devoirs?
			F	avez fait 10 séries à la télé

Past or Present?

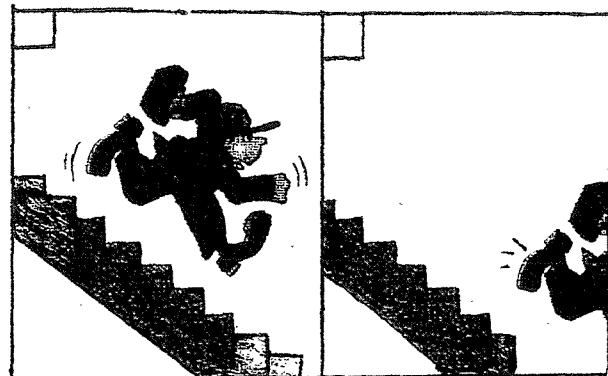
Is each sentence describing something *being* done now or something that has already been done?

The first picture in each pair shows an action being done *now*. The second picture shows that the action *has been done in the past*.

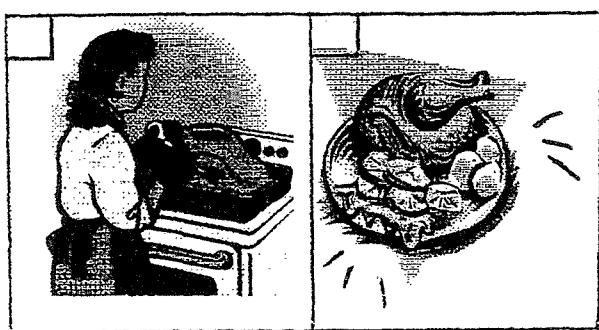
Tick the picture that best illustrates the sentence.



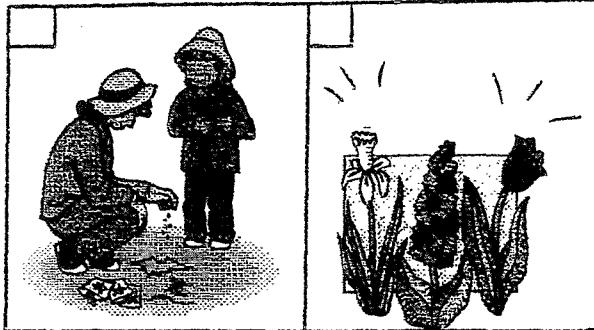
1) Nous avons écrit



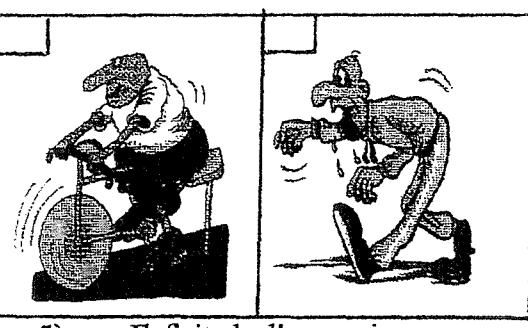
2) Il descend l'escalier



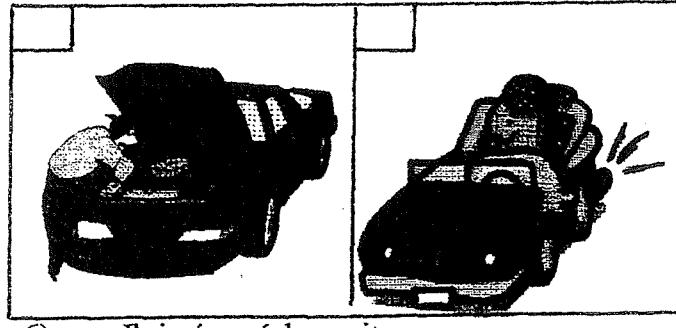
3) Elle a préparé le dîner



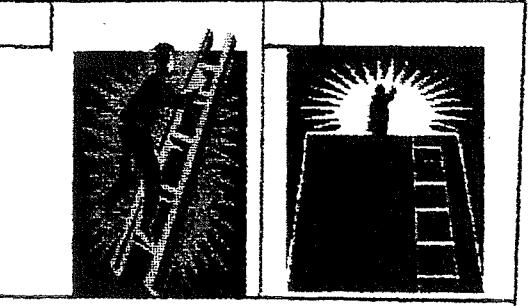
4) Elles plantent des fleurs



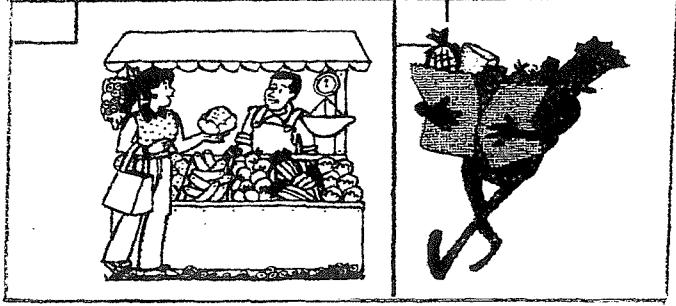
5) Il fait de l'exercice



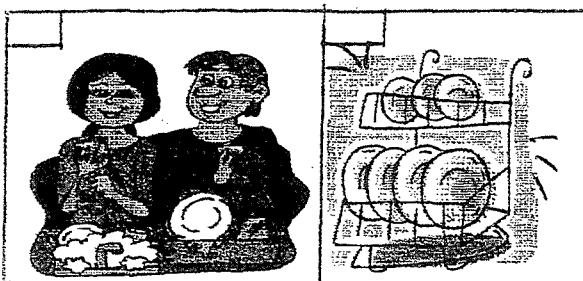
6) J'ai réparé la voiture



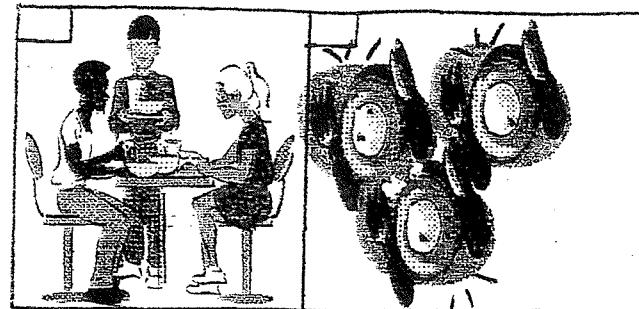
7) Il est monté l'échelle



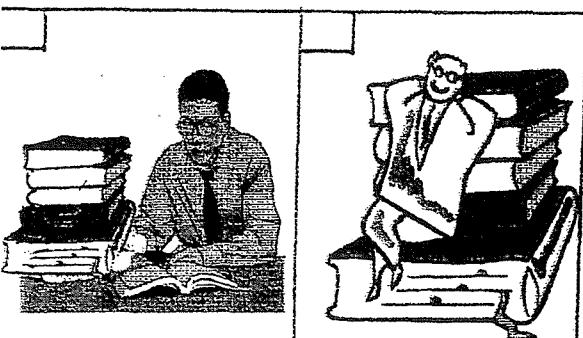
8) Tu as acheté des fruits



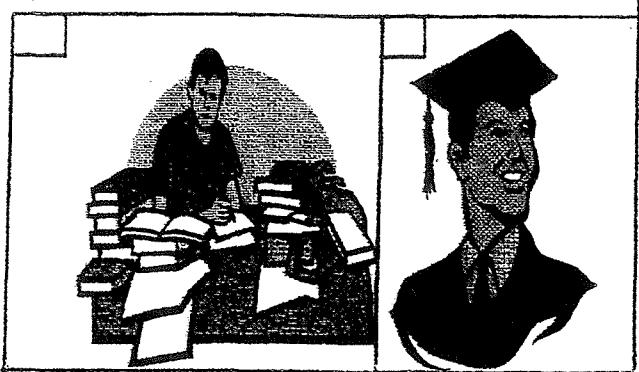
9) Ils ont lavé la vaisselle



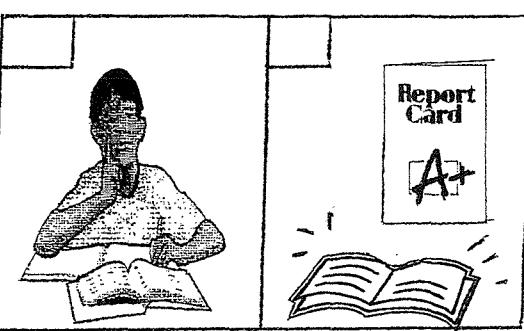
10) Nous mangeons



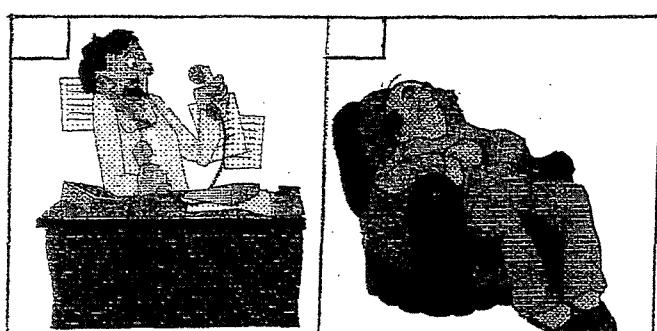
11) Vous corrigez les cahiers



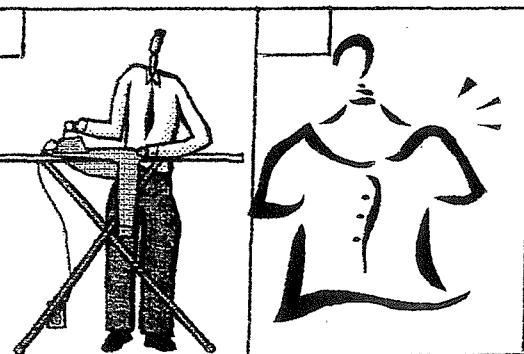
12) Vous avez étudié beaucoup



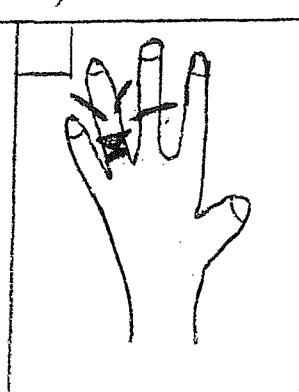
13) Il a fait les devoirs



14) Tu travailles beaucoup



15) Je repasse un chemisier



16) Ils sont descendus l'escalier



17) Il donne un anneau de mariage à Lucie

Appendix 21b The interpretation tests: listening tests

Transcript: Person & number, listening test 1

Part 1 - present

- 1) _____ commence les activités
- 2) _____ rangeons la chambre
- 3) _____ ne prépare pas le dîner
- 4) _____ ne présentez pas le projet
- 5) _____ restez à l'école
- 6) _____ ne corrigeons pas les devoirs

1, 2, 3, 4

2 null
2 ons
2 ez
1 declarative & 1 negative for each

Part 2: Perfect tense

1. _____ avons travaillé dans le jardin
2. _____ n' ont pas réservé une table
3. _____ a recyclé les bouteilles
4. _____ ne sont pas arrivés en train
5. _____ ai donné les cadeaux à ~~ma soeur~~ Paul
6. _____ ont fini les devoirs
7. _____ n' a pas acheté les fruits
8. _____ n'avons pas écouté la radio
9. _____ sont allés à 8 heures
10. _____ n'ai pas préparé le dîner

2, 3, 4

2 ai
2 a
2 avons
2 ont
2 sont
1 declarative & negative in each

rising intonation for each taken out Jan 6 2002
not testing any 'est' in perfect tense - too similar to 'ai' ???

Transcript - Present irregular - number-Test 1

- 1) [Le train] va à Paris
- 2) [La fille] a une piscine
- 3) [Elles] font les activités
- 4) [Les professeurs] sont intelligents
- 5) [Les garçons] vont à Calais
- 6) [Il] fait les devoirs
- 7) [La maison] est grande
- 8) [Les villes] ont un parc

11/15

Listening – Interpreting tense test 1

transcript

1. Je nage dans la piscine
2. Les filles ont travaillé dans la salle de classe
3. Luc a réservé une chambre
4. Je fais les activités
5. J'ai fait les devoirs
6. Elle ne range pas le salon
7. Nous avons joué au seash basket
8. Je n'ai pas dansé à l'hôtel
9. Les grands-parents sont rentrés à 5 heures
10. Marc réserve une table
11. Les filles travaillent dans la bibliothèque
12. Vous avez parlé au professeur
13. Il mange les frites
14. Elle est resté à la maison
15. J'ai nagé dans la mer
16. Je ne danse pas dan la disco
17. Il a mangé les pommes
18. Il dit bonjour au professeur
19. Nous jouons au tennis
20. Tu fermes la porte
21. Tu as fermé la fenêtre
22. Vous parlez au téléphone
23. Il a dit au revoir au médecin
24. Elle n'a pas rangé la chambre

a = 1 4 6 10 11 13 16 18 19 20 22

b = 2 3 5 7 8 9 12 14 15 17 21 23 24

Listening

Part 1

The first word of each sentence you hear has been missed out.

Which word should be at the start of each sentence? (The sentences refer to the present).

Tick ONE of: il (he) nous (we) vous (you) not sure

1)	il	nous	vous	not sure
2)	il	nous	vous	not sure
3)	il	nous	vous	not sure
4)	il	nous	vous	not sure
5)	il	nous	vous	not sure
6)	il	nous	vous	not sure

Part 2

The first word of each sentence you hear has been missed out.

Which word should be at the start of each sentence? (The sentences refer to the past).

Tick ONE of: je (I) il (he) nous (we) ils (they) not sure

1)	je	il	nous	ils	not sure
2)	je	il	nous	ils	not sure
3)	je	il	nous	ils	not sure
4)	je	il	nous	ils	not sure
5)	je	il	nous	ils	not sure
6)	je	il	nous	ils	not sure
7)	je	il	nous	ils	not sure
8)	je	il	nous	ils	not sure
9)	je	il	nous	ils	not sure
10)	je	il	nous	ils	not sure

Listening

The first word of each sentence you hear has been missed out.

Decide whether the person is talking about **ONE** person (or thing) or **MORE THAN ONE**.

Tick the word that **must** have come at the start of each sentence you hear.

Tick 'not sure' if you don't know.

1) Le garçon...

Les garçons...

Not sure

2) La ville...

Les villes...

Not sure

3) Il...

Ils...

Not sure

4) La maison...

Les maisons...

Not sure

5) Le train...

Les trains...

Not sure

6) Elle...

Elles...

Not sure

7) Le professeur...

Les professeurs...

Not sure

8) La fille...

Les filles...

Not sure

Listening now or in the past?

Indicate whether these sentences are using the present tense (i.e. talking about now) or the perfect tense (i.e. talking about the past).

Tick one column: 'present', 'past' or 'not sure'.

	present	past	not sure
1)			
2)			
3)			
4)			
5)			
6)			
7)			
8)			
9)			
10)			
11)			
12)			
13)			
14)			
15)			
16)			
17)			
18)			
19)			
20)			
21)			
22)			
23)			
24)			

Appendix 22a The production tests: writing tests

Writing

Fill in the blanks to make complete sentences.

All the sentences are about things that always happen or happen regularly.

Normalement...

- 1 Vous _____ à l'école à 8 heures en France (start)
- 2 Le supermarché _____ à 8 heures les jeudis (shut)
- 3 Normalement, il _____ la voiture le weekend (wash)
- 4 Nous _____ le pizza - c'est délicieux! (love)
- 5 Je _____ dans la piscine tous les weekends (swim)
- 6 Mon père _____ à la maison à 7 heures tous les jours (arrive)
- 7 Elles _____ les CDs les samedis (listen)
- 8 Hélène _____ la télé tous les soirs (watch)
- 9 Tu _____ le chien dans le parc (walk)

Part 2

Fill in the blanks.

All the sentences are about things that always happen or happen regularly.

Normalement...

1. Il _____ les exercices dans les cahiers [do]
2. Elles _____ à Brighton en train [go]
3. La ville _____ moderne [be]
4. Elle _____ à Paris en avion [go]
5. Ils _____ petits [be]
6. Le cinéma _____ 12 écrans (=screens) [have]
7. Les filles _____ les devoirs dans la bibliothèque [do]
8. Les ordinateurs _____ des problèmes [have]

Writing

Un Weekend Typique de Ma Mère

Write 5 sentences describing a typical weekend of your Mum (or your Dad, your Auntie etc)

You **must** include:

2 sentences about things she does alone (*She*)

2 sentences about things s/he does with other people (not with you!) (*They....*)

You must use a different verb for each sentence.

Normalement, le weekend, ma mère....

Ma mère et une amie (?)...

Mon Weekend Typique

Imagine you are your French teacher!

Write 5 sentences describing what you do on a typical weekend (yes - you may have to use your imagination about what she does, but tell it as though it was really true e.g *I... We....*)

You **must** include:

2 sentences about things you do **alone** (**without your family or friends!**) (*I...*)

2 sentences about things you do with your family and/or friends (*We...*).

You must use a different verb for each sentence.

Normalement, le weekend, je...

Nous...

Writing - what happened?

Complete the following sentences with the correct words to say what happened.

All the sentences refer to events **in the past** (last year, yesterday, last weekend etc):

- 1) L'année dernière, les hommes _____ au basket (play)
- 2) Hier, ma grandmère _____ 10 cigarettes (smoke)
- 3) Le weekend dernier, je / j' _____ dans un groupe (sing)
- 4) Tu _____ les plans la semaine dernière? (present)
- 5) L'année dernière, il _____ dans un show à Londres (dance)
- 6) Qu'est-ce que tu _____ chez tes grandparents à Noël? (eat)
- 7) Pendant les vacances, il _____ au Canada (go)
- 8) Moi et mes amis, nous _____ le dîner hier soir (prepare)
- 9) Samedi dernier, Hélène _____ dans la banque à 3 heures (enter)
- 10) Hier soir, ils _____ un film au cinéma (watch)
- 11) Le weekend dernier, ils _____ à 7 heures (arrive)

Writing - what happened?

Les Vacances de Mon Ami

Write 6 sentences describing what a friend and his (or her) family did in the holidays (you may have to guess what they did, but tell it as though it was really true e.g S/he....They ...)

You **must** include:

3 sentences about things your friend did **alone** (*s/he...*)

3 sentences about things your friend did with their family **together** (*they...*)

You must use a different verb for each sentence.

Remember!! you are writing about things that happened in the past.

Pendant les vacances, mon ami...

Ils...

Mes Vacances

Imagine you are Victoria or David Beckham!

Write 6 sentences describing the holidays you and your family had (yes - you will have to use your imagination, but tell it as though it was really true e.g I..., we...)

You **must** include:

3 sentences about things you did **by yourself** (*I...*)

3 sentences about things you did **with** your family **together** (*we...*)

You must use a different verb for every sentence.

Remember!! you are writing about things that happened in the past.

Pendant les vacances, je...

Nous...

Appendix 12 cont'd

Quick revision list for writing activity

to start = commencer

to shut = fermer

to wash = laver

to love = adorer

to swim = nager

to arrive = arriver

to listen = écouter

to watch = regarder

to walk = promener

to do = faire

to go = aller

to be = être

to have = avoir

Quick revision list for writing activity

to play = jouer

to smoke = fumer

to sing = chanter

to present = présenter

to dance = danser

to eat = manger

to go = aller

to prepare = préparer

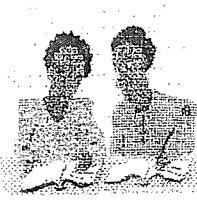
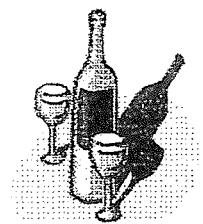
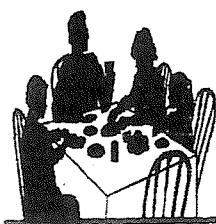
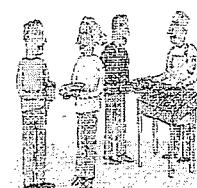
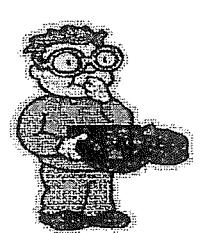
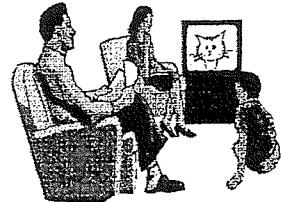
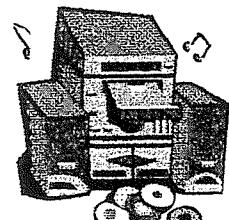
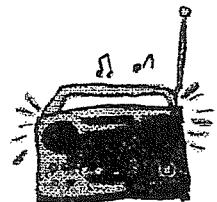
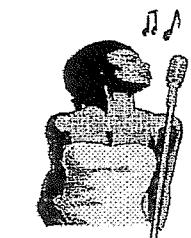
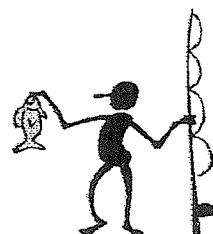
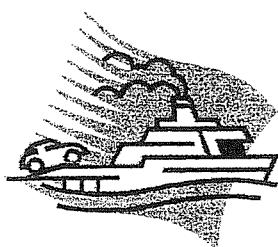
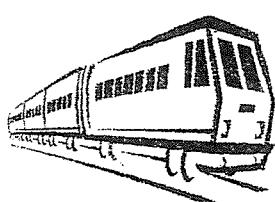
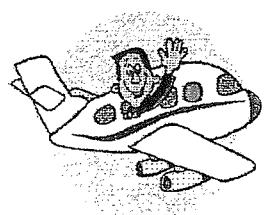
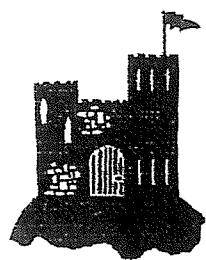
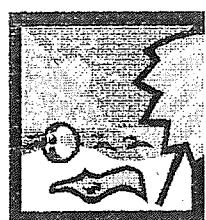
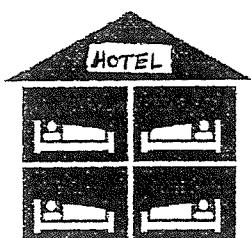
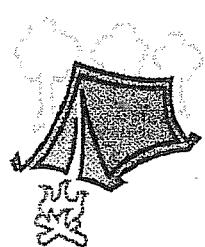
to enter = entrer

to watch = regarder

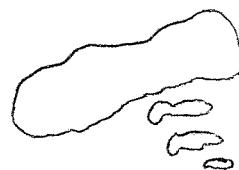
to arrive = arriver

WHAT TO SAY?? - these are JUST IDEAS to jog your memory!

(you do not have to use this sheet)



KEY TO SYMBOLS



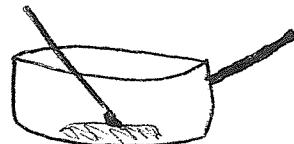
= think about (penser)



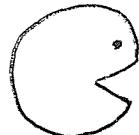
= like (aimer)



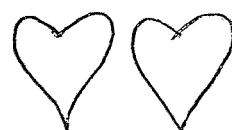
= look at (regarder)



= prepare (préparer)



= eat (manger)



= love (adorer)



= hate (détester)



= wash (laver)

EXAMPLE Speaking

You will hear an example of this activity then you will do a similar one yourself.

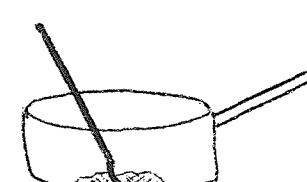
These sentences are all about chocolate mousse !

Say the whole of each sentence, replacing the symbol with the correct word.

1. Nous  à la mousse au chocolat

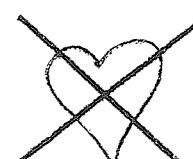
2. Vous  la mousse

3. Je  les instructions

4. Julie  le chocolat

5. Tu  la mousse

6. Ils  la mousse

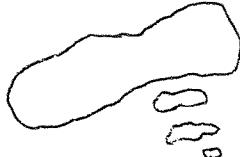
7. Les amis  la mousse

8. Elle  la vaiselle

Speaking

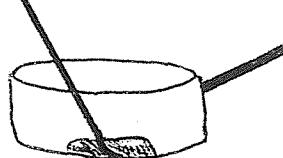
These sentences are all about chocolate mousse !

Say the whole of each sentence, replacing the symbol with the correct word.

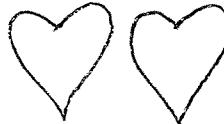
1. Elle  à la mousse au chocolat

2. Les amis  la mousse

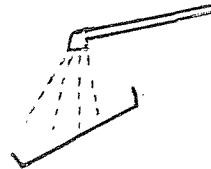
3. Marc  les instructions

4. Nous  le chocolat

5. Je  la mousse

6. Vous  la mousse

7. Ils  la mousse

8. Tu  la vaiselle

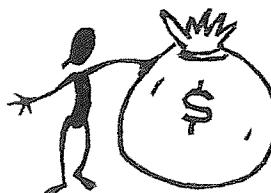
Speaking

Read out the full sentences filling in the blank with a word to make the sentence complete.

The meaning is also given in the picture just to make sure you know what you are supposed to be saying!

The sentences only need either: **DO, GO, HAVE or BE (is/are).**

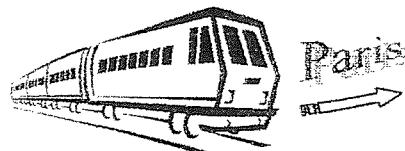
1) Il _____ riche



2) Elle _____ une idée



3) Le train _____ à Paris



4) Paul _____ de l'exercice



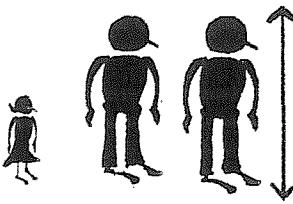
5) Les filles _____ les devoirs



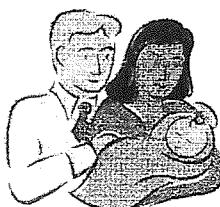
6) Ils _____ en ville



7) Les garçons _____ grands



8) Ils _____ un bébé



Revision of Symbols



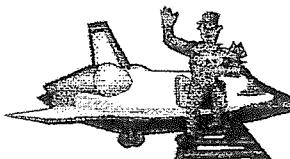
to work (travailler)



to listen (écouter)



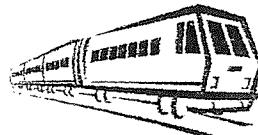
to dance (danser)



to arrive (arriver)



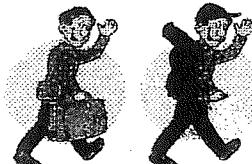
to wash (laver)



to go (aller)



to play (jouer)



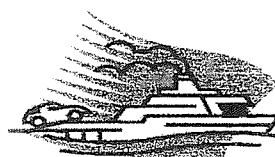
to go (aller)



to eat (manger)



to watch (regarder)



to return (rentrer)

Speaking - What happened during the holidays?

Read out the full sentences filling in the blank with a word to make the sentence complete.

The meaning is also given in the picture just to make sure you know what you are supposed to be saying! All the sentences are describing what people did during the LAST holidays.

Pendant les vacances...

1) Mon père _____ dans le jardin (work)



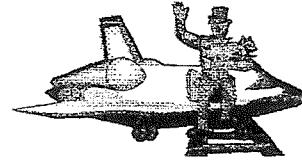
2) Vendredi soir, nous _____ la radio



3) Mes parents _____ samedi dernier



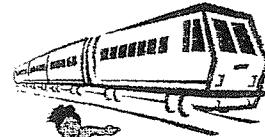
4) Il _____ en avion samedi dernier (arrive)



5) Je _____ le chien le weekend dernier



6) Ma mère _____ à Paris en train l'été dernier (go)



7) Elle _____ au tennis le weekend dernier



8) Mes frères _____ en vacances en août (go)



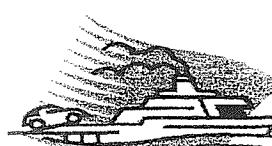
9) Tu _____ chez MacDonalds pendant les vacances?



10) Ils _____ la télé hier soir



11) Ils _____ en ferry en juillet (return)



Researcher's Instructions - Speaking

Use test 1 with the first pupil, test 2 with the second pupil, test 1 with the third pupil etc

Intro

Greet the pupil:

"Bonjour, je m'appelle Ros., assieds-toi, s'il te plaît. You have about 4 tasks to do. It will probably take about 10 minutes in total. I will explain each task to you as we go along. OK, the first task."

If they ask vocab, give them nouns but no verbs

Task 1 - Picture Narration - Mon Weekend Typique

- Make sure you have the relevant 'boy' or 'girl' tape ready and present tense task sheet 1 - the picture narration task. Boy tape for a boy, girl tape for a girl!

- Give them the present tense Task Sheet 1 - the picture narration task

Read out loud while they are reading the instructions:

"You will hear this person talking about what he (or she - if pupil is a girl!) and his (her) family normally do at the weekend. He (or she) is describing a typical weekend. When you have heard the recording, you must imagine you are this boy (or girl). Using the pictures, describe your typical weekend to the researcher, imagining you are the boy (or girl) in the first picture."

- Tell the pupil to put the earphones on, in English. Press play - make sure the pupil is looking at the pictures. The pupil will press stop.
- Remind them, in English: "Now, imagine you are that boy (or girl). Describe what you & your family do" Then ask them straight away "Normalement, le weekend qu'est-ce que tu fais et que fait ta famille?"

Make sure they say something for every picture - if they miss one out, and you see they are clearly going to, cut in and say "et que fait-il / ton frère / tes parents etc". If they have missed one out and you couldn't cut in at the time, wait until they have got to the last picture and then ask them about the one/s they missed out.

- Say "très bien, merci beaucoup, now task 2" to every pupil

Example of narration for discourse speaking present test 1 - BOY

Le samedi matin, je nage, au centre sportif. Mon père travaille dans le jardin presque tous les weekends.

Pendant l'après midi, ma soeur parle ~~à ses~~^{aux} amis au téléphone.
Mes parents promènent le chien avec mes deux frères

Le samedi soir nous mangeons chez nous ou quelquefois dans un restaurant.

Ma mère lave la vaisselle,
Et nous regardons la télé.

Je sais que plus tard mes grands-parents dansent dans le salon, ~~ils écoutent la musique et~~
~~ils dansent!~~

Press STOP

Example of narration for speaking present test 1 - GIRL

Le samedi matin, mon frère nage, au centre sportif. Mon père travaille dans le jardin presque tous les weekends.

Pendant l'après midi, je parle ~~à mes~~^{aux} amis au téléphone.
Mes parents promènent le chien avec mes deux frères

Le samedi soir nous mangeons chez nous ou quelquefois dans un restaurant.

Ma mère lave la vaisselle,
Et nous regardons la télé.

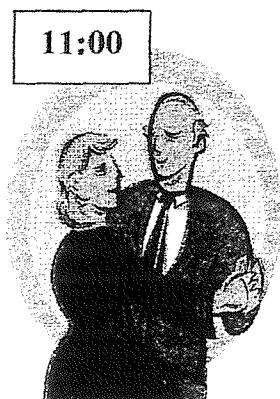
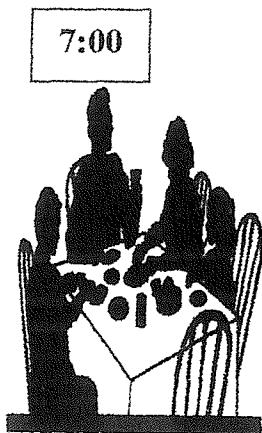
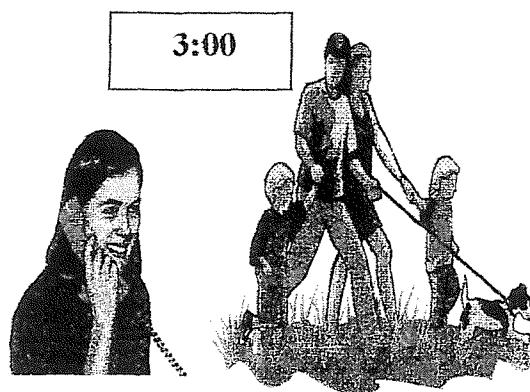
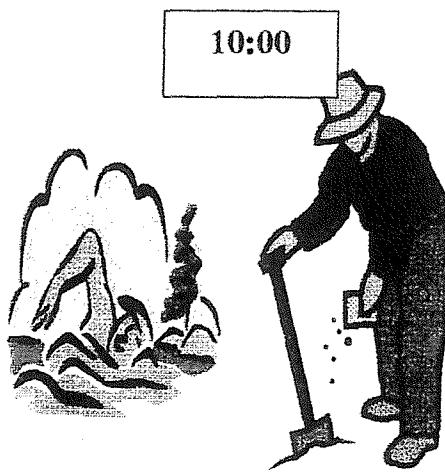
Je sais que plus tard mes grands-parents dansent dans le salon, ~~ils écoutent la musique et~~
~~ils dansent!~~

Press STOP

Speaking Task Sheet 1.

- First you will hear this person talking about what he and his family normally do at the weekend. He is describing a typical weekend.
- When you have heard the recording, you must imagine the pictures are about what you and your family normally do ON A TYPICAL SATURDAY. You are the boy swimming in the first picture. You can decide who the other pictures are about.

Describe what happens in French



Task 2 - Guided Conversation about Typical Weekend

- Give them Task Sheet 2. Show them the first part, covering up the other parts with the card. Let them read it and then say
"OK, you begin. When we have finished this part, move the card down to see part 2. You move the card down as we finish each part."

Part 1: The pupil asks you what you normally do at the weekend - say ONLY "pas beaucoup, je reste à la maison, je travaille un peu"

Part 2: They ask you about specific activities, you say "oui, je..." or "non, je ne..." (it doesn't matter which)

Part 3: They ask what you & your husband do.

Say ONLY: "nous faisons des promenades quelquefois et nous rendons visite chez des amis"

Part 4: They ask you about specific activities, you say "oui, nous..." or "non, nous ne..." (it doesn't matter which)

Part 5: Ask the pupil, "et toi, qu'est-ce que tu fais normalement le weekend?" They are asked on the sheet to say at least 2 things.

- If they don't, prompt by pointing - saying nothing - back at the pictures, raising eyebrows questioningly!
- If still no response, ask a Yes / No question - tu aides tes parents? tu ranges la maison? tu fais du sport? tu aimes la musique? - trying to get them to develop their responses using their own verb

Part 6: Ask the pupil, "et toi et ta famille qu'est-ce que vous faites normalement?" Slight pause, "Ou toi et tes amis?", (they are asked to say at least 2 things, in total)

- Say "super, merci. Now task 3"

TASK SHEET 2¹

Part 1) Ask the lady what she normally does at the weekend (Normalement, le weekend, ...)

use 'tu' for 'you'

Part 2) Find out if she

- a) washes the car **(use 'tu' for 'you')**
- b) eats in a restaurant

Part 3) Ask what she & her husband normally do at the weekend (Normalement, le weekend, ...)

you must use 'vous' for 'you'

Part 4) Find out if they

- a) watch television **(you must use 'vous' for 'you')**
- b) play tennis

Part 5) When you are asked - say at least 2 things ! (je...). Then move on to part 6 ↓

Part 6) When you are asked - say at least 2 things ! (nous...)

Task 3 - Christmas Photos

- Give them Task Sheet 3

Read the instructions out loud while they read them:

"1) You will hear somebody talking about what she and her relatives did in the holidays - imagine she is a friend of yours telling you what they did. Follow the photos as she is talking. When she finishes, press stop.

2) Then describe what they did on each day. Use the photos to help you. The researcher does not know what your friend did. You should say something about every photo.

3) Begin, "le 24 décembre..."

and use "Son père...il...elle...Alice...ils...le chien...etc"

- Tell the pupil to put the earphones on, in English. Press play - make sure the pupil is looking at the pictures. The pupil will press stop.
- Remind the pupil in English: "Now tell me what your friends did in their holidays at Christmas. Qu'est-ce qu'ils ont fait tes amis pendant les vacances...le 24 décembre?"

Don't let them look back at the task sheet while they talk - so they can't use the pronoun prompts!

Make sure they try to say something for each picture. If they are really stuck, after about 10-15 seconds (?!), ask Yes/No question which is obviously wrong- elles ont joué au football/ fait les devoirs / écrit des lettres? If still no response, ask a Yes/No question that is correct!

- Say "OK, excellent, now the last task"

Task Sheet 3 - Les Vacances de Noël

- 1) You will hear somebody talking about what she and her relatives did in the holidays - imagine she is a friend of yours telling you what they did. Follow the photos as she is talking. When she finishes, press stop.
- 2) Then describe what they did on each day. Use the photos to help you. The researcher does not know what your friend did. You should say something about every photo.
- 3) Begin, "le 24 décembre..."
and use "Son père...il...elle...Alice...ils...le chien...etc"

Tape Transcript
Narration for Christmas Photos

037

pendant les vacances

~~Nous sommes allés chez mes parents pour Noël. Nous sommes~~ restés à la maison.

Le 24 décembre mon père a regardé un match de football à la télé - ~~il est fan de Coventry City.~~

Le soir, mon frère, mon mari et mon père ont joué à "Pictionary". Ils ont joué beaucoup.

TOURNE LA PAGE

Le jour de Noël, le 25 décembre, nous avons mangé un grand dîner.

L'après midi ma nièce, qui s'appelle Alice, a joué avec les cadeaux et elle a fait du vélo.

TOURNE LA PAGE

Le 26 décembre, ~~mes parents~~ ont promené le chien. Le chien a joué dans la neige.

Le soir elles ont lu un livre. Après ça, elles sont allées au lit.

Press STOP

015

029

044

058

le 24 décembre



et le soir...



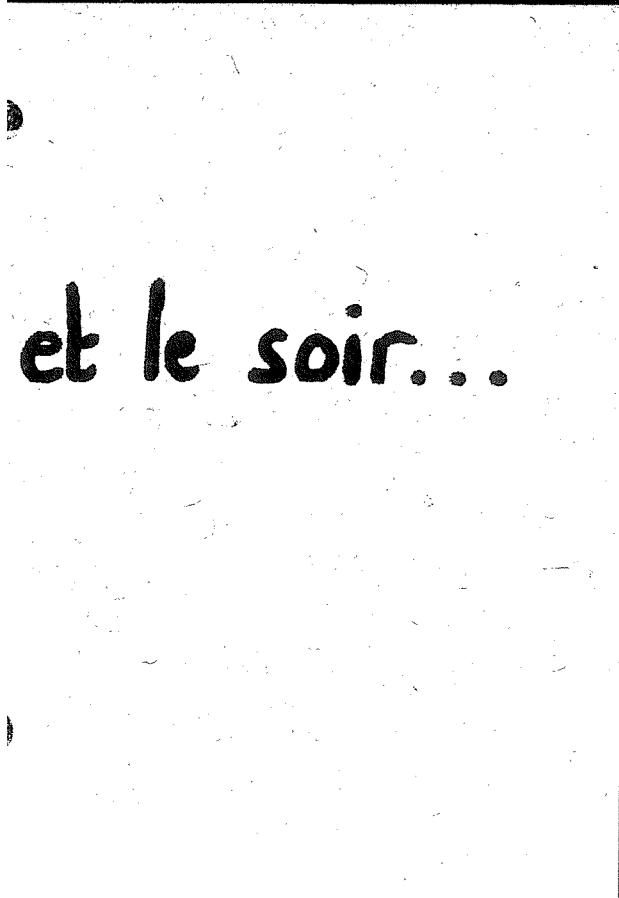
le 25 décembre



et l'après-midi...



le 26 décembre



et le soir...

Task 4 - Guided Conversation about the Holidays

- Give them Task Sheet 4. Show them the first part, covering up the other parts with the card. Let them read it and then say

"OK, you begin. When we have finished this part, move the card down to see part 2. You move the card down as we finish each part."

Part 1: The pupil asks you what you did during the holidays - say ONLY: "je suis allé chez mon frère, je n'ai pas travaillé beaucoup, c'était bien"

Part 2: They ask you about specific activities, you say "oui, j'ai..." or "non, je n'ai ..." (it doesn't matter which)

Part 3: They ask what you and your family did - say ONLY: "nous avons bu beaucoup de vin, on s'est détendu, rien de spéciale"

Part 4: They ask you about specific activities, you say "oui, nous avons..." or "non, nous n'avons pas..." (it doesn't matter which)

Part 5: Ask the pupil, "et toi, qu'est-ce que tu as fait pendant les vacances?" They are asked on the sheet to say at least 3 things.

-If they don't, prompt them simply by pointing - saying nothing - back at some of the photos, just raising eyebrows questioningly!

-If they still don't, after 10-15 seconds, then ask:

qu'est-ce que tu as mangé / joué / regardé à la télé/
tu es resté à la maison/ allé au cinéma?

-If still don't produce own language (e.g. what film they watched) ask a more direct Yes/No question!

Part 6: Ask the pupil, "et toi et ta famille, qu'est-ce que vous avez fait pendant les vacances?" Slight pause, "Ou toi et tes amis?", (they are asked to say at least 3 things, in total- if they don't, do the same as above but with 'vous').

- Say "super, merci beaucoup. C'est fini. You can go back to your class now and ask the teacher for the next student"

!!!!Change to the other test (1 or 2) for the next pupil

& check whether boy or girl as they arrive!!!!

TASK SHEET 4

1) Ask the lady what she did in the holidays ("Pendant les vacances ...)

use 'tu' for 'you'

2) Find out if she

- a) played tennis **(use 'tu' for 'you')**
- b) watched television

3) Ask what she and her husband did in the holidays ("Pendant les vacances ...)

you must use 'vous' for 'you'

4) Find out if they

- a) ate chocolate **(you must use 'vous' for 'you')**
- b) listened to music

5) When asked - say at least 3 things ! (je...). Then move on to part 6



6) When asked - say at least 3 things ! (nous...)

Appendix 23 Scoring written interlanguage

a) Sentence level

The spelling of the stem verb was not scored, as there were no unrecognisable stems¹.

There were fixed totals for all learners in the sentence level tasks as almost all the learners attempted all the items using the correct verb².

Present tense

Learners were given infinitive forms in a revision list prior to the test. 2 points were available for each item, awarded as follows:

0 points: an infinitive; past participle; attempt at a past participle (any accent on an 'e' after the stem); any attempt at an auxiliary regardless of the form of any subsequent lexical verb; unsuitable semantic choice, regardless of target-likeness of inflection.

1 point: any attempt at present inflection, including null endings; no ending (e.g. regard); non-target like inflection (e.g. mangeon for mangeons, fais for fait, allent for vont, regardez for regardons).

2 points: a correctly spelt target-like inflection.

Perfect tense

2 points were available for each obligatory context, awarded as follows:

0 points: an infinitive; an attempt at or target-like present tense; unsuitable lexical verb

0.5 points: an attempt at past participle (requiring an accent on an 'e' after stem e.g. allér, allè, regardeé) with or without an auxiliary.

0.5 points: an attempt at auxiliary if followed by a lexical verb (including *a* for *est* and *allons / et / ont / ent*).

1 point: target-like past participle (had to be a suitable lexical verb; gender & number agreements were ignored as this was not a target feature in the teaching materials).

1 point: target-like auxiliary

¹ The most potentially controversial was probably accepting promenades for promenes or promenade for promene (this was done as forms such as promenade + r / é/ ér / ez / ent were recurrent). However, if there was a determiner in front of a form of promenade, it was not counted as a verb.

² There were a few occurrences in the sentence level tasks where learners did not complete an item or used the wrong lexical verb. Analyses were carried out where such instances were eliminated and each learner had an 'individual' total possible score. However, there were no differences between this analysis and an analysis with the number of items as a fixed total for all learners. The latter is maintained throughout the thesis, as this complements the method of scoring in the narratives (which assessed learners' ability to inflect verbs they 'accessed' themselves i.e. in learner produced obligatory contexts).

b) Guided narrative level

These were scored out of a total obligatory contexts produced by each learner (see main thesis for definition). Use of L1 or L3 was not counted as an obligatory context.

Present tense

As for the sentence level tasks, except that there were a few instances of slightly ambiguous semantics where 1 point was given if the verb inflection was correct (ma soeur travaille les devoirs).

Perfect tense

As for the sentence level tasks, including that to get any score for an auxiliary, there had to be a lexical verb e.g. Nous avons en supermarché or Ils ont le fishng were given no points.

C'est, c'était, j'aime were excluded as obligatory contexts of the target features as it was considered more reliable to exclude the data rather than guess what tense / aspect the learners intended.

There were a very few occurrences of an item being inserted between the auxiliary and the lexical verb (e.g. nous sommes un visite) and 0.5 points were taken off the score.

Co-indexation to the auxiliary was allowed, as in mature grammars e.g. Nous avons joué (2 points) au tennis et lavé (2 points) le chien.

Further notes on the definition of obligatory contexts

The pronoun/subject written by the learner was used as the obligatory context if this was unambiguous (e.g. 'Je adorer le film' written after a sentence which had used the 'nous' prompt was counted as an obligatory context for 'je'). However, if learners inserted a pronoun immediately after the subject prompt then this was counted as an obligatory context for the subject prompt (not the learners' pronoun) e.g. “ma mere et son amie” [=written prompt] ‘vous avez allé’ would count as a third person plural context. If it was not clear which referent was intended for a second or third verb in a paragraph then the context was not counted e.g. ‘ma mère et une amie aller un café. Nous fait bon’ ('nous fait' was ignored). In any case, there were few such occurrences and the analyses were not at the fine-grained level of the accuracy of particular person and number inflections.

Appendix 24 Scoring oral interlanguage

Learners' final production of any verbs was scored (it was very rare that pupils had more than one attempt at any utterance). Some verbs were pronounced "orthographically" and these were considered correct e.g. mangeONS, unless this affected communication (e.g. suggesting interference from another language) in which case one mark was given if the form was likely to be correct e.g. la fille eS (for est).

The sentence level tasks were scored out of the number of items (i.e. a fixed total for all learners). The narrative tasks were scored out of the number of obligatory contexts produced by each learner (unsuitable lexical verbs and verbs with no subject were discounted from the total obligatory contexts).

Present tense

0 points: the infinitive form (unless this was [e] with vous) (it is emphasised that the same 'advantage' was experienced across all groups); chunks such as 'nous j'ai'
1 point: non-target like inflection e.g. conveying the wrong number and/or person
2 points: target-like inflection.

Perfect tense

1.5 points were available for a target-like auxiliary and 1 for a target-like past participle. A lexical verb had to be used in order to score anything for an auxiliary. These were awarded as follows:

0 points: regular 'er' infinitive e.g. je mang[e]
1 point: attempt at an auxiliary with a non-target-like past participle (e.g. 'short' lexical verb: nous a mange)
1 point: correct irregular past participle with no auxiliary e.g. il lu.
1.5 points: attempt at an auxiliary with target-like past participle e.g. nous a mang[e]
1.5 points: target-like auxiliary with non-target-like past participle e.g. il a lave, elle a lis[e]
1.5 points: a few occurrences of an intervening le/la between target-like auxiliary and past participle e.g. il a le mang[e]
2 points: target-like auxiliary and past participle.

**Appendix 25 Likert scale questions about pupils' attitudes to French,
administered just before pre, post and dp tests**

Name _____

1) Compared to other subjects, how do you like French?

Tick the phrase that you feel is most true for you:

French is my favourite subject

French is one of the subjects I usually like

French is one of the subjects I usually think is OK

French is one of my least favourite subjects

French is my least favourite subject

2) How difficult do you find French compared to other subjects?

Tick the phrase that you feel is most true for you:

French is the most difficult subject

French is one of the more difficult subjects

French is one of the subjects I usually find OK

French is one of the easier subjects

French is the easiest subject for me

Appendix 26 : Questionnaire about the intervention activities

When the teachers have split the group into two, what do you think of the activities you have been doing? Explain your thoughts as fully as possible:

1 Have you enjoyed them? Yes / No

Why / why not? Because _____

2 Did you find them helpful and useful? Yes / No

Why / why not? Because _____

3 Do you think they were any different from things you have done before? Underline:

The activities were: *a bit different* / *weren't actually that different*

If you think the activities *were* different to things you have done before,

i) how were they different? _____

ii) would you like the teacher to continue using this style of activities? Yes / No

Why / why not? _____

4 What have you been learning when you have been in the split groups? Give examples if you like.

5 What *type* of activities have you been doing and what is your opinion about them?

6 Have the activities changed the way you notice French grammar at all? Yes a bit / No not really
If so, please explain _____

7 What would you change about these activities? _____

Please write ANYTHING else you'd like to add -suggestions, comments- on the back of this sheet.

Appendix 27 Non-parametric tests

Listening tests

	Overall	Class A	Class B	Class C	Merged EI	Merged PI	Class A, EI	Class A, PI	Class B, EI	Class B, PI
N	85	28	27	30	27	28	14	14	13	14
Chi-Square	43.855	34.545	1.415	20.018	9.250	17.429	20.593	15.571	.120	3.857
df	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	.000	.000	.493	.000	.010	.000	.000	.000	.942	.145

Friedman analyses of variance of differences between listening test scores

		POST – PRE (1-tailed)	DPT – POST (2 tailed)	DPT – PRE (1 tailed)
Overall	Z	-6.309	-1.838	-6.056
	Asymp. Sig.	.000	.066	.000
Class A	Z	-4.533	-1.286	-4.579
	Asymp. Sig.	.000	.198	.000
Class B	Z	-2.326	-.216	-1.207
	Asymp. Sig.	.010	.829	.114
Class C	Z	-3.755	-2.439	-4.087
	Asymp. Sig.	.000	.015	.000
Merged EI	Z	-3.088	-1.104	-2.909
	Asymp. Sig.	.001	.270	.001
Merged PI	Z	-4.021	-.184	-3.462
	Asymp. Sig.	.000	.854	.001
Class A, EI	Z	-3.235	-2.120	-3.297
	Asymp. Sig.	.001	.034	.001
Class A, PI	Z	-3.204	-.341	-3.235
	Asymp. Sig.	.001	.733	.001
Class B, EI	Z	-.707	-.489	-.039
	Asymp. Sig.	.240	.624	.485
Class B, PI	Z	-2.262	-.220	-1.570
	Asymp. Sig.	.012	.826	.058

Wilcoxon tests of differences between pairs of listening test scores

Reading tests

	Overall	Class A	Class B	Class C	Merged EI	Merged PI	Class A, EI	Class A, PI	Class B, EI	Class B, PI
N	85	28	27	30	27	28	14	14	13	14
Chi-Square	46.447	25.327	8.712	28.359	8.766	17.290	13.857	11.593	1.059	10.302
df	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	.000	.000	.013	.000	.012	.000	.001	.003	.589	.006

Friedman analyses of variance of differences between reading test scores

		POST - PRE	DPT - POST	DPT - PRE
Overall	Z	-5.642	-3.611	-6.703
	Asymp. Sig.	.000	.000	.000
Class A	Z	-4.160	-3.067	-4.314
	Asymp. Sig.	.000	.002	.000
Class B	Z	-2.518	-1.759	-2.236
	Asymp. Sig.	.006	.079	.013
Class C	Z	-2.873	-3.343	-4.373
	Asymp. Sig.	.002	.001	.000
Merged EI	Z	-2.912	-1.229	-2.973
	Asymp. Sig.	.002	.219	.002
Merged PI	Z	-3.965	-.806	-3.991
	Asymp. Sig.	.000	.420	.000
Class A, EI	Z	-3.111	-2.106	-3.109
	Asymp. Sig.	.001	.035	.001
Class A, PI	Z	-2.764	-2.295	-3.046
	Asymp. Sig.	.003	.022	.001
Class B, EI	Z	-.456	-.880	-.235
	Asymp. Sig.	.325	.379	.407
Class B, PI	Z	-2.830	-1.646	-2.639
	Asymp. Sig.	.003	.100	.004

Wilcoxon tests of differences between pairs of reading test scores

Writing tests

	Overall	Class A	Class B	Class C	Merged EI	Merged PI	Class A, EI	Class A, PI	Class B, EI	Class B, PI
N	80	29	26	25	26	29	14	15	12	14
Chi-Square	52.900	38.345	5.846	24.000	14.846	24.069	17.714	20.800	3.500	6.143
df	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	.000	.000	.054	.000	.001	.000	.000	.000	.174	.046

Friedman analyses of variance of differences between the writing test scores

		POST - PRE	DPT - POST	DPT - PRE
Overall	Z	-4.854	-4.341	-6.824
	Asymp. Sig.	.000	.000	.000
Class A	Z	-4.508	-2.692	-4.617
	Asymp. Sig.	.000	.007	.000
Class B	Z	-2.337	-.793	-1.943
	Asymp. Sig.	.010	.428	.026
Class C	Z	-.745	-3.216	-4.418
	Asymp. Sig.	.228	.001	.000
Merged EI	Z	-3.454	-1.177	-2.984
	Asymp. Sig.	.001	.239	.002
Merged PI	Z	-3.968	-2.335	-4.271
	Asymp. Sig.	.000	.020	.000
Class A, EI	Z	-3.170	-1.915	-3.233
	Asymp. Sig.	.001	.056	.001
Class A, PI	Z	-3.351	-1.704	-3.408
	Asymp. Sig.	.001	.088	.001
Class B, EI	Z	-1.295	-.664	-.235
	Asymp. Sig.	.098	.507	.407
Class B, PI	Z	-1.915	-1.664	-2.417
	Asymp. Sig.	.028	.096	.008

Wilcoxon tests of differences between pairs of writing test scores

Speaking tests

	Overall	Class A	Class B	Class C	Merged EI	Merged PI	Class A, EI	Class A, PI	Class B, EI	Class B, PI
	48	16	17	15	15	18	8	8	7	10
Chi-Square	25.978	11.143	2.469	16.915	2.133	11.910	7.000	4.323	1.143	9.056
df	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	.000	.004	.291	.000	.344	.003	.030	.115	.565	.011

Friedman analyses of variance of differences between the speaking test scores

		POST - PRE	DPT - POST	DPT - PRE
Overall	Z	-4.170	-1.273	-4.950
	Asymp. Sig.	.000	.203	.000
Class A	Z	-2.767	-1.552	-3.069
	Asymp. Sig.	.003	.121	.001
Class B	Z	-1.448	-.796	-1.397
	Asymp. Sig.	.074	.426	0.081
Class C	Z	-2.840	-1.023	-3.409
	Asymp. Sig.	.003	.306	0.001
Merged EI	Z	-1.704	-.852	-1.905
	Asymp. Sig.	.044	.394	.029
Merged PI	Z	-2.486	.000	-2.820
	Asymp. Sig.	.007	1.000	.003
Class A, EI	Z	-2.380	-1.260	-2.313
	Asymp. Sig.	.009	.208	.011
Class A, PI	Z	-1.540	-.980	-2.028
	Asymp. Sig.	.062	.327	.022
Class B, EI	Z	-.507	.000	-.423
	Asymp. Sig.	.306	1.000	.336
Class B, PI	Z	-1.956	-1.051	-1.897
	Asymp. Sig.	.025	.293	.029

Wilcoxon tests of differences between pairs of speaking test scores

Appendix 29 Tables of the tests of the normality of the distribution of data

		Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig. ¹	Statistic	df	Sig.
LISTPRE	Overall	.103	85	.026	.969	85	.040
LISTPOST		.101	85	.031	.963	85	.014
LISTDPT		.083	85	.200	.980	85	.225
LISTPRE	Class A	.145	28	.138	.952	28	.228
	Class B	.124	27	.200	.966	27	.494
	Class C	.096	30	.200	.949	30	.162
LISTPOST	Class A	.098	28	.200	.964	28	.442
	Class B	.143	27	.168	.959	27	.356
	Class C	.128	30	.200	.983	30	.895
LISTDPT	Class A	.104	28	.200	.957	28	.303
	Class B	.150	27	.120	.956	27	.298
	Class C	.084	30	.200	.971	30	.577
LISTPRE	Merged EI	.118	27	.200	.972	27	.648
	Merged PI	.145	28	.137	.940	28	.109
LISTPOST	Merged EI	.188	27	.015	.896	27	.011
	Merged PI	.116	28	.200	.947	28	.165
LISTDPT	Merged EI	.121	27	.200	.934	27	.086
	Merged PI	.132	28	.200	.970	28	.580
Class A							
LISTPRE	EI	.152	14	.200	.961	14	.746
	PI	.137	14	.200	.941	14	.433
LISTPOST	EI	.150	14	.200	.953	14	.616
	PI	.195	14	.157	.907	14	.141
LISTDPT	EI	.126	14	.200	.959	14	.701
	PI	.199	14	.139	.914	14	.177
Class B							
LISTPRE	EI	.123	13	.200	.971	13	.910
	PI	.154	14	.200	.922	14	.233
LISTPOST	EI	.168	13	.200	.965	13	.828
	PI	.106	14	.200	.948	14	.525
LISTDPT	EI	.203	13	.147	.830	13	.016
	PI	.201	14	.128	.925	14	.257

Tests of the normality of the distributions of the listening test data

¹ This is a lower bound of the true significance i.e. K-S test does not produce p values above 0.200.

		Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
READPRE	Overall	.109	85	.015	.954	85	.004
READPOST		.100	85	.036	.950	85	.002
READDPT		.114	85	.008	.932	85	.000
	GROUP						
READPRE	EI	.173	27	.038	.928	27	.062
	PI	.173	28	.031	.890	28	.007
READPOST	EI	.183	27	.021	.898	27	.012
	PI	.135	29	.191	.902	29	.011
READDPT	EI	.204	27	.005	.832	27	.001
	PI	.165	29	.041	.905	29	.013
	CLASS						
READPRE	A	.124	28	.200	.953	28	.232
	B	.179	27	.026	.943	27	.145
	C	.108	30	.200	.983	30	.901
READPOST	A	.149	28	.116	.919	28	.033
	B	.168	27	.050	.937	27	.104
	C	.108	30	.200	.973	30	.613
READDPT	A	.159	28	.069	.904	28	.014
	B	.128	27	.200	.875	27	.004
	C	.113	30	.200	.956	30	.238
	Class A						
READPRE	EI	.126	14	.200	.963	14	.768
	PI	.178	14	.200	.938	14	.396
READPOST	EI	.134	14	.200	.945	14	.491
	PI	.187	14	.200	.897	14	.103
READDPT	EI	.258	14	.012	.811	14	.007
	PI	.161	14	.200	.938	14	.389
	Class B						
READPRE	EI	.174	13	.200	.926	13	.302
	PI	.207	14	.108	.934	14	.350
READPOST	EI	.244	13	.033	.894	13	.112
	PI	.161	14	.200	.932	14	.329
READDPT	EI	.286	13	.005	.672	13	.000
	PI	.096	14	.200	.984	14	.991

Tests of normality of the distributions of the reading totals

		Kolmogorov-Smirnov			Shapiro-Wilk		
GROUP		Statistic	df	Sig.	Statistic	df	Sig.
SREADPRE	EI	.148	27	.134	.959	27	.360
	PI	.154	28	.088	.948	28	.178
SREADPOS	EI	.173	27	.036	.917	27	.034
	PI	.118	28	.200	.918	28	.031
SREADDPT	EI	.192	27	.012	.838	27	.001
	PI	.145	28	.139	.937	28	.095
CLASS							
SREADPRE	A	.112	28	.200	.972	28	.624
	B	.151	27	.118	.972	27	.660
	C	.111	30	.200	.977	30	.733
SREADPOS	A	.154	28	.089	.893	28	.008
	B	.144	27	.160	.962	27	.402
	C	.102	30	.200	.977	30	.749
SREADDPT	A	.185	28	.015	.853	28	.001
	B	.132	27	.200	.926	27	.054
	C	.112	30	.200	.952	30	.186
CLASS 1							
GROUP							
SREADPRE	1.00	.134	14	.200	.963	14	.774
	2.00	.145	14	.200	.973	14	.909
SREADPOS	1.00	.159	14	.200	.913	14	.176
	2.00	.187	14	.200	.882	14	.062
SREADDPT	1.00	.281	14	.004	.766	14	.002
	2.00	.159	14	.200	.894	14	.092
CLASS 2							
GROUP							
SREADPRE	1.00	.143	13	.200	.962	13	.782
	2.00	.182	14	.200	.939	14	.402
SREADPOS	1.00	.216	13	.100	.940	13	.453
	2.00	.153	14	.200	.943	14	.457
SREADDPT	1.00	.252	13	.023	.748	13	.002
	2.00	.113	14	.200	.981	14	.980

Tests of the normality of the distributions of the square root of the reading data

		Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
WRITPRE	Overall	.176	80	.000	.833	80	.000
WRITPOST	Overall	.168	80	.000	.920	80	.000
WRITDPT	Overall	.095	80	.071	.929	80	.000
WRITPRE	Class A	.182	29	.015	.900	29	.010
	Class B	.121	26	.200	.905	26	.020
	Class C	.194	25	.016	.880	25	.007
WRITPOST	Class A	.095	29	.200	.961	29	.351
	Class B	.147	26	.155	.933	26	.090
	Class C	.134	25	.200	.928	25	.080
WRITDPT	Class A	.093	29	.200	.954	29	.226
	Class B	.140	26	.200	.890	26	.009
	Class C	.158	25	.107	.927	25	.073
WRITPRE	Merged EI	.205	26	.006	.797	26	.000
	Merged PI	.173	29	.027	.825	29	.000
WRITPOST	Merged EI	.195	26	.012	.893	26	.011
	Merged PI	.188	29	.010	.853	29	.001
WRITDPT	Merged EI	.203	26	.007	.880	26	.006
	Merged PI	.141	29	.146	.904	29	.012
Class A							
WRITPRE	EI	.215	14	.079	.882	14	.061
	PI	.182	15	.195	.881	15	.050
WRITPOST	EI	.147	14	.200	.968	14	.848
	PI	.163	15	.200	.899	15	.093
WRITDPT	EI	.193	14	.168	.950	14	.564
	PI	.145	15	.200	.942	15	.407
Class B							
WRITPRE	EI	.136	12	.200	.950	12	.631
	PI	.273	14	.006	.770	14	.002
WRITPOST	EI	.226	12	.093	.951	12	.654
	PI	.227	14	.048	.864	14	.035
WRITDPT	EI	.203	12	.184	.919	12	.281
	PI	.180	14	.200	.913	14	.173

Tests of the normality of the distributions of writing test data

		Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Overall PRE		.185	80	.000	.878	80	.000
Overall POST		.152	80	.000	.940	80	.001
Overall DPT		.150	80	.000	.950	80	.003
PRE	Class A	.272	29	.000	.814	29	.000
	Class B	.213	26	.004	.888	26	.009
	Class C	.213	25	.005	.869	25	.004
POST	Class A	.241	29	.000	.810	29	.000
	Class B	.176	26	.038	.955	26	.295
	Class C	.242	25	.001	.890	25	.011
DPT	Class A	.151	29	.091	.962	29	.374
	Class B	.163	26	.075	.935	26	.100
	Class C	.202	25	.010	.884	25	.008
PRE	EI	.302	26	.000	.771	26	.000
	PI	.195	29	.006	.880	29	.003
POST	EI	.225	26	.002	.881	26	.006
	PI	.108	29	.200	.959	29	.312
DPT	EI	.187	26	.020	.918	26	.040
	PI	.187	29	.011	.944	29	.131
Class A, EI							
PRE		.309	14	.001	.731	14	.001
POST		.301	14	.001	.696	14	.000
DPT		.158	14	.200	.923	14	.245
Class B, EI							
PRE		.280	12	.010	.852	12	.038
POST		.173	12	.200	.963	12	.825
DPT		.236	12	.063	.893	12	.128
Class A, PI							
PRE		.204	15	.093	.899	15	.090
POST		.252	15	.011	.812	15	.005
DPT		.154	15	.200 ²	.949	15	.505
Class B, P							
PRE		.185	14	.200	.915	14	.185
POST		.225	14	.053	.934	14	.348
DPT		.214	14	.083	.900	14	.112

Tests of the normality of the distributions of obligatory contexts produced in written guided written narrative tasks

² 0.200 is the maximum p value provided by the K-S test in SPSS - it gives the maximum indication available the samples can not be considered normally distributed.

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Overall						
SRWRIPRE	.083	80	.200	.970	80	.056
SRWRIPOS	.106	80	.026	.971	80	.068
SRWRITDP	.081	80	.200	.955	80	.007
CLASS A						
SRWRIPRE	.107	29	.200	.986	29	.963
SRWRIPOS	.105	29	.200	.971	29	.577
SRWRITDP	.129	29	.200	.926	29	.044
CLASS B						
SRWRIPRE	.082	26	.200	.976	26	.770
SRWRIPOS	.155	26	.111	.919	26	.043
SRWRITDP	.106	26	.200	.960	26	.383
Class C						
SRWRIPRE	.144	25	.194	.934	25	.108
SRWRIPOS	.138	25	.200	.934	25	.106
SRWRITDP	.154	25	.127	.941	25	.155
Merged EI						
SRWRIPRE	.127	26	.200	.917	26	.039
SRWRIPOS	.139	26	.200	.944	26	.172
SRWRITDP	.161	26	.083	.920	26	.044
Merged PI group						
SRWRIPRE	.094	29	.200	.971	29	.595
SRWRIPOS	.107	29	.200	.962	29	.366
SRWRITDP	.085	29	.200	.961	29	.351
Class A, EI group						
SRWRIPRE	.146	14	.200	.942	14	.447
SRWRIPOS	.209	14	.100	.921	14	.225
SRWRITDP	.231	14	.041	.899	14	.110
Class B, EI group						
SRWRIPRE	.137	12	.200	.963	12	.825
SRWRIPOS	.183	12	.200	.951	12	.652
SRWRITDP	.194	12	.200	.950	12	.630
Class A, PI group						
SRWRIPRE	.152	15	.200	.965	15	.775
SRWRIPOS	.131	15	.200	.950	15	.531
SRWRITDP	.139	15	.200	.938	15	.362
Class B, PI group						
SRWRIPRE	.157	14	.200	.934	14	.350
SRWRIPOS	.244	14	.023	.857	14	.027
SRWRITDP	.125	14	.200	.957	14	.673

Normality of distribution tests for transformed (square root) writing scores

		Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
SPKPRE	Overall	.210	48	.000	.758	48	.000
SPKPOST	Overall	.162	48	.003	.881	48	.000
SPKDPT	Overall	.149	48	.009	.903	48	.001
SPKPRE	Class A	.287	16	.001	.683	16	.000
	Class B	.113	17	.200	.945	17	.387
	Class C	.239	15	.021	.839	15	.012
SPKPOST	Class A	.172	16	.200	.914	16	.133
	Class B	.181	17	.143	.871	17	.023
	Class C	.193	15	.137	.894	15	.077
SPKDPT	Class A	.127	16	.200	.933	16	.267
	Class B	.170	17	.200	.920	17	.146
	Class C	.234	15	.027	.931	15	.285
SPKPRE	Merged EI	.186	15	.173	.932	15	.297
	Merged PI	.307	18	.000	.589	18	.000
SPKPOST	Merged EI	.191	15	.144	.895	15	.080
	Merged PI	.230	18	.013	.828	18	.004
SPKDPT	Merged EI	.150	15	.200	.907	15	.121
	Merged PI	.247	18	.005	.773	18	.001
Class A							
SPKPRE	EI	.304	8	.028	.890	8	.232
	PI	.351	8	.004	.575	8	.000
SPKPOST	EI	.121	8	.200	.964	8	.849
	PI	.341	8	.007	.782	8	.018
SPKDPT	EI	.179	8	.200	.953	8	.737
	PI	.272	8	.085	.784	8	.019
Class B							
SPKPRE	EI	.204	7	.200	.905	7	.359
	PI	.168	10	.200	.936	10	.514
SPKPOST	EI	.172	7	.200	.973	7	.919
	PI	.192	10	.200	.939	10	.538
SPKDPT	EI	.265	7	.146	.864	7	.166
	PI	.164	10	.200	.957	10	.747

Normality of distribution tests for speaking scores

		Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
SQRTSKPR	Overall	.152	48	.007	.906	48	.001
SQRTSKPO	Overall	.122	48	.071	.960	48	.097
SQRTSKDP	Overall	.088	48	.200	.956	48	.070
SQRTSKPR	Class A	.236	16	.018	.802	16	.003
	Class B	.126	17	.200	.938	17	.291
	Class C	.180	15	.200	.940	15	.378
SQRTSKPO	Class A	.133	16	.200	.975	16	.916
	Class B	.140	17	.200	.949	17	.438
	Class C	.142	15	.200	.950	15	.522
SQRTSKDP	Class A	.117	16	.200	.951	16	.510
	Class B	.165	17	.200	.916	17	.124
	Class C	.193	15	.136	.956	15	.620
SQRTSKPR	Merged EI	.140	15	.200	.977	15	.942
	Merged PI	.224	18	.018	.765	18	.001
SQRTSKPO	Merged EI	.171	15	.200	.930	15	.271
	Merged PI	.162	18	.200	.934	18	.224
SQRTSKDP	Merged EI	.139	15	.200	.928	15	.255
	Merged PI	.180	18	.126	.886	18	.033
Class A							
SQRTSKPR	EI	.266	8	.099	.939	8	.598
	PI	.319	8	.016	.655	8	.001
SQRTSKPO	EI	.151	8	.200	.941	8	.620
	PI	.286	8	.053	.893	8	.248
SQRTSKDP	EI	.208	8	.200	.938	8	.589
	PI	.248	8	.157	.833	8	.064
Class B							
SQRTSKPR	EI	.190	7	.200	.923	7	.492
	PI	.184	10	.200	.916	10	.329
SQRTSKPO	EI	.197	7	.200	.971	7	.903
	PI	.147	10	.200	.975	10	.932
SQRTSKDP	EI	.247	7	.200	.878	7	.216
	PI	.184	10	.200	.936	10	.511

Normality of distribution tests for transformed (square root) of the speaking scores

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Overall						
TOTCON PRE	.132	24	.200	.955	24	.345
TOTCON POST	.122	24	.200	.960	24	.433
TOTCON DPT	.213	24	.006	.933	24	.114
CLASS						
TOTCON PRE	A	.168	8	.200	.935	8
	B	.340	8	.007	.838	8
	C	.233	8	.200	.881	8
TOTCON POST	A	.178	8	.200	.873	8
	B	.170	8	.200	.983	8
	C	.189	8	.200	.943	8
TOTCON DP	A	.219	8	.200	.936	8
	B	.153	8	.200	.952	8
	C	.233	8	.200	.881	8
GROUP						
TOTCON PRE	EI	.162	8	.200	.932	8
	PI	.296	8	.037	.860	8
TOTCON POST	EI	.179	8	.200	.952	8
	PI	.149	8	.200	.923	8
TOTCON DPT	EI	.248	8	.161	.913	8
	PI	.215	8	.200	.857	8
Class A, EI						
TOTCON PRE		.307	4	.	.729	4
TOTCON POST		.210	4	.	.982	4
TOTCON DP		.162	4	.	.989	4
Class A, PI						
TOTCON PRE		.250	4	.	.963	4
TOTCON POST		.283	4	.	.863	4
TOTCON DP		.269	4	.	.844	4
Class B, EI						
TOTCON PRE		.210	4	.	.982	4
TOTCON POST		.214	4	.	.963	4
TOTCON DP		.307	4	.	.729	4
Class B, PI						
TOTCON PRE		.415	4	.	.716	4
TOTCON POST		.151	4	.	.993	4
TOTCON DP		.229	4	.	.895	4

Normality of the distribution tests for the obligatory contexts produced in narrative / conversation speaking tests

Appendix 30 Interview with teachers A & B regarding intervention period

First I'll ask you about the actual teaching - then we can talk about the administration of the study and any other issues.

- 1) What have you felt about the materials we have been using for the split group teaching?

Blue materials & green - show examples

- 2) Do you think they are any different from materials you were using before?
If so how?

- 3) What do you think the POINT of this activity was? Show green referential and affective activities? And blue comparisons for each of these.

- 4) What do you think about the level of the materials?

Easy / hard / too many linguistic features / not enough? (compared with the rate you would have got through the linguistic features (present and perfect) normally?)

- 5) Motivational value of the materials (both blue and referential and affective activities?) Interesting materials / fun or boring?

- 6) How do you think the pupils reacted? Both in terms of

- a) motivational value? - compared to previous materials...?
 - b) And linguistically?

- 7) Average score would you say the class should get on average on 10 questions asking them to write verbs in present? In perfect?

- 8) When you were using the materials, did you prefer one or another? - which?
Why?

Did this affect how you taught with the materials?

- 9) What did you feel were the advantages of one or another? - did you think one set would be better for anything? Make a more marked improvement in any area?

- 10) What improvements do you think could be made to the materials?

- 11) Have you thought about adapting any of the ideas for your own teaching?
Could the ideas behind the materials be useful for other classes / languages / linguistic features?

- 12) Do you think you will use any of the materials again?

Appendix 31 Descriptive statistics of the language test scores

Sample		pre	post	dp
Class A	PI	52.98 17.35 (14)	72.08 15.94 (15)	69.72 20.00 (15)
	EI	50.30 16.78 (14)	64.73 15.88 (14)	72.47 14.11 (14)
Class B	PI	40.33 9.16 (14)	49.11 11.08 (14)	48.96 15.69 (14)
	EI	41.51 11.039 (13)	43.59 5.80 (13)	41.51 14.35 (13)
Merged PI		46.65 15.06 (28)	60.99 17.90 (29)	59.70 20.63 (29)
Merged EI		46.06 14.74 (27)	54.55 16.04 (27)	57.56 21.05 (27)
Class A		51.64 16.80 (28)	68.53 16.06 (29)	71.05 17.16 (29)
Class B		40.90 9.93 (27)	46.45 9.21 (27)	45.37 15.25 (27)
Class C		49.72 12.62 (30)	59.31 13.87 (30)	64.93 13.80 (30)
Overall		47.55 14.06 (85)	58.38 16.04 (86)	60.85 18.72 (86)

Total listening mean scores, standard deviations (and n) in each grouping sample

Sample		pre	post	dp
Class A	PI	48.83 17.81 (14)	61.94 20.86 (15)	68.85 22.43 (15)
	EI	52.73 17.69 (14)	64.16 19.70 (14)	68.83 18.89 (14)
Class B	PI	25.32 6.55 (14)	36.75 9.44 (14)	34.16 7.31 (14)
	EI	26.85 9.16 (13)	28.53 8.95 (13)	27.83 13.10 (13)
PI		37.08 17.80 (28)	49.78 20.57 (29)	52.10 24.24 (29)
EI		40.27 19.20 (27)	47.00 23.67 (27)	49.09 26.33 (27)
Class A		50.78 17.53 (28)	63.01 19.98 (29)	68.84 20.43 (29)
Class B		26.06 7.80 (27)	32.79 9.95 (27)	31.11 10.78 (27)
Class C		51.27 18.16 (30)	56.67 15.66 (30)	64.79 19.16 (30)
Overall		43.10 19.19 (85)	51.31 20.28 (86)	55.58 24.05 (86)

Total reading mean scores, standard deviations (and n) in each grouping sample

Sample		pre	post	dp
Class A	PI	24.08	42.79	49.99
		<i>18.50</i>	<i>26.08</i>	<i>25.59</i>
		(15)	(15)	(15)
	EI	21.88	45.44	51.91
		<i>17.24</i>	<i>22.09</i>	<i>22.13</i>
		(14)	(14)	(14)
Class B	PI	6.80	13.47	17.22
		<i>8.10</i>	<i>8.96</i>	<i>12.30</i>
		(14)	(14)	(14)
	EI	10.14	13.02	13.37
		<i>5.06</i>	<i>6.18</i>	<i>14.63</i>
		12	13	13
PI		15.74	28.64	34.17
		16.70	24.49	25.99
		29	29	29
EI		16.46	29.83	33.36
		14.19	23.11	26.99
		26	27	27
Class A		23.01	44.07	50.92
		17.62	23.84	23.57
		29	29	29
Class B		8.34	13.25	15.37
		6.95	7.60	13.35
		26	27	27
Class C		33.20	38.25	51.08
		23.82	23.10	24.18
		30	27	28
Overall		22.12	32.15	39.55
		20.41	23.70	26.72
		85	83	84

Total writing mean scores, standard deviations (and n) in each grouping sample

	Overall	N	Mean	Std. Deviation
Overall	SPKPRE	49	19.6808	13.77704
	SPKPOST	49	27.1045	17.24942
	SPKDP	48	29.6567	18.79982
Class A	SPKPRE	16	19.9693	15.59924
	SPKPOST	16	29.0973	18.72817
	SPKDP	16	34.6065	21.02597
Class B	SPKPRE	17	14.8778	6.16261
	SPKPOST	17	19.2135	10.90442
	SPKDP	17	17.6762	8.27738
Class C	SPKPRE	16	24.4955	16.50988
	SPKPOST	16	33.4959	18.87271
	SPKDP	15	37.9547	18.93377
Merged EI	SPKPRE	15	17.4737	8.75540
	SPKPOST	15	23.8252	13.21355
	SPKDP	15	27.0518	16.98719
Merged PI	SPKPRE	18	17.2403	14.13738
	SPKPOST	18	24.1561	18.02545
	SPKDP	18	24.9123	18.80681
Class A, EI	SPKPRE	8	19.8326	9.69548
	SPKPOST	8	32.3160	12.63574
	SPKDP	8	38.8396	14.08727
Class A, PI	SPKPRE	8	20.1059	20.67341
	SPKPOST	8	25.8786	23.83805
	SPKDP	8	30.3734	26.60695
Class B, EI	SPKPRE	7	14.7779	7.30064
	SPKPOST	7	14.1214	3.89108
	SPKDP	7	13.5801	6.69230
Class B, PI	SPKPRE	10	14.9477	5.65420
	SPKPOST	10	22.7780	12.92325
	SPKDP	10	20.5434	8.35241

Total speaking mean scores, standard deviations and n in each grouping sample

Appendix 32 Levene's tests of equality of variances

For variances to be considered homogenous at the 95% confidence level, sig. must be more than 0.05; at the 99% confidence level (increasing the chance of accepting the null hypothesis of 'no difference'), sig. must be more than 0.01 (this is a less conservative outcome for the purposes of finding equal variances).

Measure	Between subject variables	Repeated measure	F	df1	df2	Sig.
Listening	Classes & groups	Pre	2.182	4	80	.078
		Post	3.635	4	80	.009
		DP	2.260	4	80	.070
	Class A	Pre	.082	1	26	.777
		Post	.094	1	26	.761
		DP	5.505	1	26	.027
	Class B	Pre	.125	1	25	.727
		Post	7.202	1	25	.013
		DP	.785	1	25	.384
	Classes	Pre	4.442	2	82	.015
		Post	4.410	2	83	.015
		DP	.830	2	83	.439
	Groups	Pre	.372	2	82	.690
		Post	1.218	2	83	.301
		DP	5.079	2	83	.008
	Classes A & B	Pre	2.757	3	51	.052
		Post	5.885	3	51	.002
		DP	2.505	3	51	.069
	Groups	Pre- post gains	.287	2	82	.751
		Post – dp gains	1.099	2	83	.338
		Pre – dp gains	.398	2	82	.673
	Classes	Pre- post gains	.141	2	82	.869
		Post – dp gains	1.158	2	83	.319
		Pre – dp gains	2.028	2	82	.138
Reading	Classes & groups	Pre	3.541	4	80	.010
		Post	4.605	4	80	.002
		DP	4.403	4	80	.003
	Class A	Pre	.000	1	26	.986
		Post	.360	1	26	.554
		DP	.815	1	26	.375
	Class B	Pre	1.195	1	25	.285
		Post	.544	1	25	.468
		DP	.933	1	25	.343
	Classes	Pre	7.124	2	82	.001
		Post	6.879	2	83	.002
		DP	6.308	2	83	.003

	Groups	Pre	.395	2	82	.675
		Post	5.531	2	83	.006
		DP	4.789	2	83	.011
	Classes A & B	Pre	3.657	3	51	.018
		Post	6.543	3	51	.001
		DP	4.346	3	51	.008
	Classes A & B	$\sqrt{\text{Pre}}$	1.727	3	51	.173
		$\sqrt{\text{Post}}$	3.218	3	51	.030
		$\sqrt{\text{DP}}$	2.186	3	51	.101
	Groups	Pre- post gains	1.316	2	82	.274
		Post - dp gains	1.157	2	83	.319
		Pre - dp gains	1.989	2	82	.143
	Classes	Pre- post gains	.527	2	82	.592
		Post - dp gains	3.970	2	83	.023
		Pre - dp gains	.782	2	82	.461
Writing	Classes & groups	Pre	7.976	4	75	.000
		Post	7.206	4	75	.000
		DP	6.002	4	75	.000
	Classes & groups	$\sqrt{\text{Pre}}$	3.065	4	75	.021
		$\sqrt{\text{Post}}$	3.124	4	75	.020
		$\sqrt{\text{DP}}$	1.102	4	75	.362
	Class A	Pre	.003	1	27	.958
		Post	.243	1	27	.626
		DP	.562	1	27	.460
	Class B	Pre	2.302	1	24	.142
		Post	4.763	1	24	.039
		DP	4.637	1	24	.042
	Classes	Pre	13.337	2	82	.000
		Post	15.694	2	80	.000
		DP	7.312	2	81	.001
	Groups	Pre	5.569	2	82	.005
		Post	.223	2	80	.800
		DP	.611	2	81	.545
	Classes A & B	Pre	4.413	3	51	.008
		Post	6.201	3	51	.001
		DP	6.918	3	51	.001
	Classes	Pre- post gains	3.731	2	79	.028
		Post - dp gains	.348	2	78	.707
		Pre - dp gains	4.266	2	80	.017
	Groups	Pre- post gains	.289	2	79	.750
		Post - dp gains	.088	2	78	.916
		Pre - dp gains	3.296	2	80	.042
Speaking	Classes & groups	Pre	1.438	4	43	.238
		Post	2.569	4	43	.051
		DP	4.335	4	43	.005

	Class A	Pre	1.082	1	14	.316
		Post	1.486	1	14	.243
		DP	3.164	1	14	.097
	Class B	Pre	.526	1	15	.480
		Post	4.492	1	15	.051
		DP	.254	1	15	.621
	Classes	Pre	2.185	2	46	.124
		Post	2.289	2	46	.113
		DP	5.520	2	46	.007
	Groups	Pre	1.146	2	46	.327
		Post	.369	2	46	.694
		DP	.217	2	45	.806
	Classes A & B	Pre	1.662	3	29	.197
		Post	2.963	3	29	.048
		DP	6.206	3	29	.002
	Classes	Pre- post gains	.098	2	46	.907
		Post - dp gains	2.466	2	46	.096
		Pre - dp gains	2.645	2	46	.082
	Groups	Pre- post gains	.289	2	79	.750
		Post - dp gains	.088	2	78	.916
		Pre - dp gains	3.296	2	80	.042

Appendix 33 Analyses without class C: repeated measures with class and group as between group factors, and planned contrasts

The following variables were found to have unequal error variances, according to Levene's test (see appendix 32): Listening: the post test data's variances; Reading: all (the transformed scores (square root) provided a dataset with equal variances for the pre and dp test, though only at 99% confidence level for the post test - analysis of the transformed scores is presented here); Writing: all; Speaking: post test (borderline), dp These analyses are therefore presented as indications of trends only. Appendix 27 provides the statistics for non-parametric analyses (which, by their nature, include analyses without class C).

Source	Type III Sum of Squares	df	Mean Square	F	Sig. ¹
LIST	4925.606	2	2462.803	27.707	.000
LIST * GROUP	285.360	2	142.680	1.605	.206
LIST * CLASS	1721.868	2	860.934	9.686	.000
LIST * GROUP * CLASS	435.766	2	217.883	2.451	.091

Source	LIST	Type III Sum of Squares	df	Mean Square	F	Sig.
LIST	Post vs. Pre	7074.246	1	7074.246	54.456	.000
	Dp vs. Pre	7689.738	1	7689.738	37.374	.000
	Post vs. Dp	12.835	1	12.835	.065	.800
LIST * GROUP	Post vs. Pre	523.794	1	523.794	4.032	.050
	Dp vs. Pre	30.368	1	30.368	.148	.702
	Post vs. Dp	301.919	1	301.919	1.527	.222
LIST * CLASS	Post vs. Pre	1922.388	1	1922.388	14.798	.000
	Dp vs. Pre	3102.642	1	3102.642	15.080	.000
	Post vs. Dp	140.572	1	140.572	.711	.403
LIST * GROUP * CLASS	Post vs. Pre	3.728	1	3.728	.029	.866
	Dp vs. Pre	701.040	1	701.040	3.407	.071
	Post vs. Dp	602.530	1	602.530	3.048	.087

¹ Mauchly's Test of Sphericity showed that this assumption was upheld (0.927, 3.800, 2, p=0.150).

Source	Type III Sum of Squares	df	Mean Square	F	Sig. ²
READ	3928.111	1.687	2328.238	38.485	.000
READ * GROUP	261.488	1.687	154.987	2.562	.092
READ * CLASS	1091.007	1.687	646.653	10.689	.000
READ * GROUP * CLASS	131.759	1.687	78.095	1.291	.277

Source	READ	Type III Sum of Squares	df	Mean Square	F	Sig.
READ	Post vs. Pre	4702.128	1	4702.128	42.858	.000
	Dp vs. Pre	6876.251	1	6876.251	50.662	.000
	Post vs. dp	205.954	1	205.954	3.389	.071
READ * GROUP	Post vs. Pre	399.918	1	399.918	3.645	.062
	Dp vs. Pre	384.393	1	384.393	2.832	.099
	Post vs. dp	.154	1	.154	.003	.960
READ * CLASS	Post vs. Pre	399.812	1	399.812	3.644	.062
	Dp vs. Pre	2167.638	1	2167.638	15.970	.000
	Post vs. dp	705.571	1	705.571	11.612	.001
READ * GROUP * CLASS	Post vs. Pre	260.784	1	260.784	2.377	.129
	Dp vs. Pre	90.375	1	90.375	.666	.418
	Post vs. dp	44.119	1	44.119	.726	.398

Source	Type III Sum of Squares	df	Mean Square	F	Sig. ³
$\sqrt{\text{READ}}$	19.987	1.617	12.360	32.510	.000
$\sqrt{\text{READ} * \text{GROUP}}$	1.966	1.617	1.216	3.199	.056
$\sqrt{\text{READ} * \text{CLASS}}$	3.413	1.617	2.111	5.552	.009
$\sqrt{\text{READ} * \text{GROUP} * \text{CLASS}}$	1.149	1.617	.711	1.870	.168

² Mauchly's test of sphericity showed that this assumption was violated, so the Greenhouse-Geisser adjustment was used for calculating the p values for the ANOVA (Field 2000 p334). Mauchly's W.815 Approx. Chi-Square10.254, 2, p=.006, Greenhouse-Geisser .844)

³ Mauchly's test of sphericity showed that this assumption was not upheld (0.763, 13.516, 2 p=0.001). The Greenhouse Geisser (0.808) is given to correct for this (without this adjustment the ANOVA produced a statistically significant interaction p=0.45, and the Huynh-Feldt adjustment produced a p value of 0.052).

Source	$\sqrt{\text{READ}}$	Type III Sum of Squares	df	Mean Square	F	Sig.
$\sqrt{\text{READ}}$	Post vs. Pre	26.003	1	26.003	37.396	.000
	Dp vs. Pre	33.485	1	33.485	40.634	.000
	Post vs. Dp	.472	1	.472	1.454	.233
$\sqrt{\text{READ}} * \text{GROUP}$	Post vs. Pre	2.949	1	2.949	4.241	.045
	Dp vs. Pre	2.951	1	2.951	3.580	.064
	Post vs. Dp	2.618E-07	1	2.618E-07	.000	.999
$\sqrt{\text{READ}} * \text{CLASS}$	Post vs. Pre	.448	1	.448	.645	.426
	Dp vs. Pre	6.361	1	6.361	7.719	.008
	Post vs. Dp	3.431	1	3.431	10.559	.002
$\sqrt{\text{READ}} * \text{GROUP} * \text{CLASS}$	Post vs. Pre	2.114	1	2.114	3.040	.087
	Dp vs. Pre	1.209	1	1.209	1.467	.231
	Post vs. Dp	.126	1	.126	.387	.537

Source	Type III Sum of Squares	df	Mean Square	F	Sig. ⁴
WRIT	8230.423	1.715	4798.494	42.556	.000
WRIT * CLASS	3822.932	1.715	2228.843	19.767	.000
WRIT * GROUP	119.183	1.715	69.486	.616	.518
WRIT * CLASS * GROUP	385.024	1.715	224.476	1.991	.149

Source	WRIT	Type III Sum of Squares	df	Mean Square	F	Sig.
WRIT	Post vs. Pre	9140.489	1	9140.489	44.330	.000
	Dp vs. Pre	14859.437	1	14859.437	58.420	.000
	Post vs. Dp	691.342	1	691.342	5.777	.020
WRIT * CLASS	Post vs. Pre	3679.154	1	3679.154	17.843	.000
	Dp vs. Pre	7203.236	1	7203.236	28.320	.000
	Post vs. Dp	586.407	1	586.407	4.900	.031*
WRIT * GROUP	Post vs. Pre	3.238	1	3.238	.016	.901
	Dp vs. Pre	153.260	1	153.260	.603	.441
	Post vs. Dp	201.051	1	201.051	1.680	.201
WRIT * CLASS * GROUP	Post vs. Pre	259.486	1	259.486	1.258	.267
	Dp vs. Pre	763.012	1	763.012	3.000	.089
	Post vs. Dp	132.574	1	132.574	1.108	.298

⁴ Mauchly's Test of Sphericity showed that this assumption was not upheld, so the Greenhouse-Geisser correction was used for the calculation of p (Mauchly's W.834, Approx. Chi-Square9.078, 2, p=.011 Greenhouse-Geisser.858)

* This is slightly different from the analysis with class C (p=0.050)

Source	Type III Sum of Squares	df	Mean Square	F	Sig. ⁵
SPK	1250.158	2	625.079	11.191	.000
SPK * CLASS	630.301	2	315.151	5.642	.006
SPK * GROUP	14.065	2	7.032	.126	.882
SPK * CLASS * GROUP	319.465	2	159.733	2.860	.065 [#]

Source	SPK	Type III Sum of Squares	df	Mean Square	F	Sig.
SPK	Post vs. Pre	1312.099	1	1312.099	13.775	.001
	Dp vs. Pre	2300.527	1	2300.527	18.664	.000
	Post vs. Dp	137.849	1	137.849	1.182	.286
SPK * CLASS	Post vs. Pre	249.190	1	249.190	2.616	.117
	Dp vs. Pre	1255.627	1	1255.627	10.187	.003
	Post vs. Dp	386.086	1	386.086	3.310	.079
SPK * GROUP	Post vs. Pre	6.400	1	6.400	.067	.797
	Dp vs. Pre	7.684	1	7.684	.062	.805
	Post vs. Dp	28.109	1	28.109	.241	.627
SPK * CLASS * GROUP	Post vs. Pre	468.626	1	468.626	4.920	.035
	Dp vs. Pre	489.542	1	489.542	3.972	.056
	Post vs. Dp	.228	1	.228	.002	.965

⁵ Mauchly's Test of Sphericity showed that no adjustment was needed (Mauchley's W 0.977, approx. Chi-square 0.648, 2, p=0.723).

[#] The main difference that this analysis has with the previous analysis is that the interaction spktest*class*group is no longer significant at the 95% confidence level. However, it is approaching significance, suggesting that there is likely to be some tendency for the type of instruction of have a differential impact depending on the class learners are in

Appendix 34 Statistics for the analysis of gains between pre, post and dp tests: merged groups and groups in classes A and B

Comparing the gains made by different instructional groups

One way ANOVAs and multiple post hoc comparisons using Gabriel's adjustment (as the sample sizes were unequal) were carried out to find the source of any differences found between the gains made by different instructional groups. Equal variances could be assumed for all but the pre-dp test gains in the reading, writing, and speaking variables (see appendix 32 for the results of Levene's test for the equality of variances). Non-parametric tests supported the findings here (appendix 27).

Gains		Sum of Squares	df	Mean Square	F	Sig.
Pre – post	Between Groups	558.466	2	279.233	1.878	.159
	Within Groups	12191.144	82	148.672		
	Total	12749.610	84			
Post – dp	Between Groups	717.370	2	358.685	1.988	.143
	Within Groups	14974.051	83	180.410		
	Total	15691.421	85			
Pre – dp	Between Groups	210.278	2	105.139	.429	.653
	Within Groups	20111.925	82	245.267		
	Total	20322.203	84			

One way ANOVA; Listening tests; merged PI, EI and class C

Gains		Sum of Squares	df	Mean Square	F	Sig.
Pre – post	Between Groups	683.678	2	341.839	3.212	.045
	Within Groups	8726.137	82	106.416		
	Total	9409.815	84			
Post – dp	Between Groups	683.784	2	341.892	3.894	.024
	Within Groups	7288.282	83	87.811		
	Total	7972.066	85			
Pre – dp	Between Groups	434.088	2	217.044	1.458	.239
	Within Groups	12209.177	82	148.892		
	Total	12643.265	84			

One way ANOVA; Reading tests; merged PI, EI and class C

Gains	(I) GROUP	(J) GROUP	Mean Diff(I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower	Upper
Pre – post	PI	EI	5.2152	2.78243	.180	-1.5615	11.9920
		Class C	6.5553	2.71068	.052	-.0460	13.1565
	Class C	EI	-1.3400	2.73652	.947	-8.0029	5.3229
Post – dp	PI	EI	.2311	2.50603	1.000	-5.8702	6.3325
		Class C	-5.8017	2.44028	.058	-11.7437	.1403
	Class C	EI	6.0328	2.48581	.051	-.0182	12.0838

Post hoc tests; Reading tests; merged PI, EI and class C

Gains		Sum of Squares	df	Mean Square	F	Sig.
Pre – post	Between Groups	1461.058	2	730.529	2.755	.070
	Within Groups	20951.246	79	265.206		
	Total	22412.304	81			
Post – dp	Between Groups	1284.754	2	642.377	3.750	.028
	Within Groups	13359.961	78	171.282		
	Total	14644.714	80			
Pre – dp	Between Groups	83.615	2	41.808	.129	.879
	Within Groups	25867.101	80	323.339		
	Total	25950.716	82			

One way ANOVA; Writing tests; merged PI, EI and class C

Gains		Mean Diff (I-J)	Std. Error	Sig.	95% Confidence Interval		
	(I) GROUP	(J) GROUP			Lower	Upper	
Post-dp	Merged PI	Merged EI	2.0059	3.50000	.918	-6.5262	10.5381
	Class C	Merged EI	9.4695	3.63250	.032	.6146	18.3244
	Class C	Merged PI	7.4636	3.57177	.114	-1.2389	16.1661

Post hoc tests; Writing tests; merged PI, EI and class C

Gains		Sum of Squares	df	Mean Square	F	Sig.
Pre – post	Between Groups	1461.058	2	730.529	2.755	.070
	Within Groups	20951.246	79	265.206		
	Total	22412.304	81			
Post – dp	Between Groups	1284.754	2	642.377	3.750	.028
	Within Groups	13359.961	78	171.282		
	Total	14644.714	80			
Pre – dp	Between Groups	83.615	2	41.808	.129	.879
	Within Groups	25867.101	80	323.339		
	Total	25950.716	82			

One way ANOVA; Speaking tests; merged PI, EI and class C

Gains			Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Dependent Variable	(I) GROUP	(J) GROUP					
Pre-post	EI	Class C	9.5014	4.58053	.106	-1.5610	20.5638
	PI	EI	-1.0780	4.47111	.968	-11.8734	9.7175
	PI	Class C	8.4234	4.21906	.123	-1.7469	18.5937
post-dp	EI	Class C	-9.4695	3.80851	.043	-18.6775	-.2615
	PI	EI	2.0059	3.34737	.821	-6.0665	10.0784
	PI	Class C	-7.4636	3.63305	.111	-16.2563	1.3291

Post hoc tests; Speaking tests; merged PI, EI and class C

Comparing the gains made by the different instructional groups in classes A and B

The gains made by the two instructional groups within each of classes A and B were analysed using independent samples t tests. The t tests are shown as two tailed tests, though p values for one tailed tests can be obtained by dividing the p value provided by 2.

		Levene's Test for Equality of Variances ¹		t-test for Equality of Means							
		Gains	F	Sig.	t	df	Sig. (2-tailed)	Mean Diff	Std. Error	95% Confidence Difference	Interval of the Diff
<i>Listening</i>										Lower	Upper
Class A	pre - post	.170	.683	.1.306	26	.203	.-5.654	4.32971	-14.554	3.24556	
	Post - dp	.099	.756	2.086	27	.047	10.100	4.84295	.16366	20.0374	
	Pre - dp	.130	.722	1.261	26	.218	5.657	4.48604	-.3.5640	14.8783	
Class B	pre - post	.067	.798	-.1.533	25	.138	-.6.696	4.36675	-.15.689	2.29733	
	Post - dp	.847	.366	-.330	25	.744	-.1.934	5.86820	-.14.020	10.1510	
	Pre - dp	1.383	.251	-.1.356	25	.187	-.8.630	6.36656	-.21.743	4.48124	
<i>Reading</i>											
Class A	Pre-post	2.142	.155	-.266	26	.792	-.1.038	3.89825	-.9.0515	6.97441	
	Post - dp	2.033	.165	-.662	27	.513	-.2.231	3.37019	-.9.1466	4.68350	
	Pre - dp	1.059	.313	-.555	26	.583	-.2.725	4.90605	-.12.809	7.35952	
Class B	Pre-post	1.467	.237	-.2.380	25	.025	-.9.753	4.09770	-.18.192	-.1.31371	
	Post - dp	.034	.856	.803	25	.429	1.898	2.36250	-.2.96770	6.76364	
	Pre - dp	.078	.782	-.2.023	25	.054	-.7.855	3.88349	-.15.8533	.14310	
<i>Writing</i>											
Class A	Pre- post	.140	.711	.742	27	.464	4.8456	6.52667	-.8.5460	18.2371	
	Post - dp	.767	.389	-.157	27	.876	-.7211	4.59091	-.10.140	8.69862	
	Pre - dp	.127	.724	.579	27	.568	4.1244	7.12832	-.10.501	18.7505	
Class B	Pre- post	.964	.336	-.1.031	24	.313	-.3.871	3.75557	-.11.622	3.87932	
	Post - dp	.133	.719	-.702	25	.489	-.3.397	4.84262	-.13.372	6.57489	
	Pre - dp	1.030	.320	-.2.446	24	.022	-.10.82	4.42556	-.19.957	-.1.69002	
<i>Speaking</i>											
Class A	Pre- post	.309	.587	1.383	14	.188	6.7100	4.85008	-.3.69239	17.1123	
	Post - dp	.520	.483	.328	14	.748	2.0288	6.18223	-.11.230	15.2883	
	Pre - dp	.000	.994	1.229	14	.239	8.7388	7.10845	-.6.5073	23.9848	
Class B	Pre- post	.233	.636	-.1.755	15	.100	-.8.488	4.83634	-.18.796	1.81984	
	Post - dp	.125	.728	.378	15	.711	1.6960	4.48466	-.7.8628	11.2548	
	Pre - dp	.532	.477	-.1.957	15	.069	-	3.47176	-.14.192	.60731	
							6.7926				

¹ Note equal variances could be assumed

Appendix 37 Extracts of responses to the questionnaire completed after the intervention

Each extract (i.e. with a line space underneath) is from a different pupil's response.

The responses are to question 1: *Have you enjoyed the activities?*, unless otherwise stated. See appendix 26 for the questionnaire itself.

Enriched Input learners

School A

They are more fun than normal lessons but you are still learning the same and I take things in better if its fun Q4 The past tense and how you can get mixed up and miss out little words

At the beginning they give a tip on how to recognise the different words, and this has helped me a lot

It makes me look closely at the text I'm reading so I'm more aware of the little things

Q6 yes because its more focused on the difference in if its you, I, he etc

Q6 yes if a sentence already tells me its in the past: le weekend dernier, je regardé une film, I don't just notice the subject, also I learn the ending of the verb

Q6 yes I notice the il a and tu as more

Q5 Past and present endings. verbs. ...It now seeps into my head instead of going in 1 ear and out the other!

School B

I learnt more then than in the last 3 years

When we do them [the activities] time flies... We get more done and it's organised... I have learnt how to listen to French... [did they change the way you notice grammar?] only a bit, I have learnt more words than grammar

I found the listening a waste of time because nothing stretched our knowledge of French

I have learned a lot in the last couple of months than I have learnt in 2 years!

SPKCONT	Type III Sum of Squares	df	Mean Square	F	Sig.
Post vs. Pre	1020.747	1	1020.747	14.544	.001
Dp vs. Pre	784.021	1	784.021	24.143	.000
Post vs. Dp	15.593	1	15.593	.217	.646

Within subject planned contrasts comparing the number of oral obligatory contexts produced

Nonparametric tests

	Overall	Class A	Class B	Class C	Merged EI	Merged PI	Class A EI	Class A PI	Class B EI	Class B PI
N	24	8	8	8	8	8	4	4	4	4
Chi-Square	12.179	4.963	6.467	6.421	11.630	3.467	6.000	.933	5.733	2.800
df	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	.002	.084	.039	.041	.003	.177	.050	.627	.057	.247

Friedman tests to compare the production of obligatory contexts in different oral tests amongst different samples.

		POST - PRE	DP - POST	DP - PRE
Overall	Z	-3.302	-.506	-2.977
	Asymp. Sig.	.001	.613	.003
Class B	Z	-2.254	-1.101	-2.205
	Asymp. Sig.	.024	.271	.027
Class C	Z	-1.876	-1.231	-2.001
	Asymp. Sig.	.061	.254	0.049
Merged EI	Z	-2.213	-2.214	-2.371
	Asymp. Sig.	.027	.027	.018
Class A EI	Z	-1.633	-1.604	-1.604
	Asymp. Sig.	.102	.109	.109

Wilcoxon 2 tailed paired tests to compare the number of contexts produced in each test in each of the samples where the Friedman test found a statistically significant difference

Correlations between contexts produced and accuracy

As the n was small and some of the data sets were not normally distributed,

Spearman's Rho (a non parametric test) was calculated to explore whether there was any possible relationship between the number of contexts produced and the accuracy of learners' productions. Statistically significant correlations were as follows:

	Spearman's rho Spk test score with contexts produced	Sig. (2-tailed)	N
Overall, post	.584	.003	24
Class A, post	.764	.027	8
Class A, PI group, post	.949	.051	4
Merged PI group, post	.731	.040	8
Overall, dp	.559	.005	24
Merged EI group, dp	.761	.028	8

Appendix 35 Statistical procedures used to explore whether the number of contexts produced in the writing tests varied according to test, class or group

	PRE			POST			DP			
	present tense contexts	perfect tense contexts	Mean total possible ¹	present tense contexts	perfect tense contexts	Mean total possible	present tense contexts	perfect tense contexts	Mean total possible	
Class A	mean	8.17	10.72	93.79	8.83	11.10	95.86	9.00	12.03	98.07
	σ	1.67	2.00	6.20	2.54	1.70	6.91	2.38	0.73	4.77
Class B	mean	8.35	9.73	92.15	9.41	11.44	97.70	9.52	11.56	98.15
	σ	1.94	3.60	9.03	1.78	1.78	6.67	1.89	1.01	3.88
Class C	mean	8.70	10.47	94.33	9.63	11.07	97.41	9.11	12.29	98.79
	σ	2.67	2.26	8.81	1.21	2.20	6.15	1.55	1.15	4.33
overall	mean	8.41	10.33	93.05	9.28	11.20	96.96	9.20	11.96	98.33
	σ	2.13	2.67	8.96	1.95	1.89	6.56	1.96	1.01	4.31

Mean obligatory contexts and standard deviations for narrative written tasks and for total mean score available for written production score, by class

It was found that the distributions of the several of the samples cannot be considered as normally distributed at the 95% confidence level (appendix 29): Class A, EI, pre & post test; Class A, PI, post test (though it is significant at the 99% level); Class B, EI, pre test (though Mauchley's $W = .852$, sig. 0.038). Therefore non-parametric tests were used where necessary.

A repeated measures ANOVA was carried out in order to assess whether the differences between the number of contexts produced at the different tests were significant and, if so, to see if this was dependent on any of the grouping variables (class or input-type group). Levene's test of equality of error variances showed that the variances of the groups in the post and dp can be considered as equal, though the pre test variances were not at the 95% confidence level. The ANOVA is treated with caution.

	F	df1	df2	Sig.
PRE	2.762	4	75	.034
POST	.432	4	75	.785
DP	.315	4	75	.867

Levene's test of equality of error variances in the number of contexts produced in the written pre, post and dp tests

¹ i.e. once the sentence totals have been added to the number of learner generated contexts, and each of these multiplied by two (as two points were available for each item).

Source	Type III Sum of Squares	df	Mean Square	F	Sig. ²
CONTEXTS	255.195	2	127.597	15.742	.000
CONTEXTS * GROUP	26.289	2	13.145	1.622	.201
CONTEXTS * CLASS	16.359	2	8.180	1.009	.367
CONTEXTS * GROUP * CLASS	26.048	2	13.024	1.607	.204

Repeated measures ANOVA on the number of contexts produced in the written tests

The ANOVA suggested there was a statistically significant difference between the contexts produced at different tests. The within-subjects planned contrasts show that the differences were between pre and post and between pre and dp tests. There were no statistically significant differences between post and dp tests.

Source	Contexts produced	Type III Sum of Squares	df	Mean Square	F	Sig.
CONTEXTS	Post vs. Pre	259.472	1	259.472	14.050	.000
	Dp vs. Pre	474.030	1	474.030	27.168	.000
	Post vs. Dp	32.082	1	32.082	2.523	.116
CONTEXTS * GROUP	Post vs. Pre	48.324	1	48.324	2.617	.110
	Dp vs. Pre	2.854	1	2.854	.164	.687
	Post vs. Dp	27.689	1	27.689	2.177	.144
CONTEXTS * CLASS	Post vs. Pre	32.714	1	32.714	1.771	.187
	Dp vs. Pre	7.830	1	7.830	.449	.505
	Post vs. Dp	8.535	1	8.535	.671	.415
CONTEXTS * GROUP * CLASS	Post vs. Pre	52.063	1	52.063	2.819	.097
	Dp vs. Pre	14.184	1	14.184	.813	.370
	Post vs. Dp	11.898	1	11.898	.936	.337

Planned contrasts comparing the number of obligatory contexts produced at each writing test

The ANOVA showed no statistically significant effect of class or group on the number of context produced. However, the planned contrasts show that between pre and post test the p value for the statistical significance of the interaction wrcontexts*class*group would be significant at the 0.10 confidence level. This may suggest the existence of some tendency that the number of contexts produced may depend on both the class learners were in and the input-type group. As the non-normal distribution of data and the non-equal variances at pre test may affect the results of this parametric test, non-parametric tests (Friedman and Wilcoxon tests)

² Mauchly's Test of Sphericity showed that no adjustment was needed for the calculation of the p values in the ANOVA (Mauchly's W 0.952, approx Chi-Square, 3.622, 2, p=0.164).

were also carried out to compare the number of contexts produced between each test in each of the samples.

	Overall	Class A	Class B	Class C	Merged EI	Merged PI	Class A EI	Class B EI	Class A PI	Class B PI
N	80	29	26	25	26	29	14	12	15	14
Chi-Square	19.531	8.019	5.312	10.539	5.312	7.635	2.680	2.837	5.593	9.120
df	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	.000	.018	.070	.005	.070	.022	.262	.242	.061	.010

Friedman tests to compare the production of obligatory contexts in different writing tests amongst different samples.

		POST - PRE	DP - POST	DP - PRE
Overall	Z	-3.389	-1.275	-5.091
	Asymp. Sig.	.001	.202	.000
Class A	Z	-.734	-1.662	-3.395
	Asymp. Sig.	.463	.097	.001
Class C	Z	-2.825	-.550	-2.509
	Asymp. Sig.	.005	.583	.012
Merged PI	Z	-2.317	.000	-3.257
	Asymp. Sig.	.021	1.000	.001
Class A PI	Z	-.667	-1.260	-2.558
	Asymp. Sig.	.505	.208	.011
Class B PI	Z	-2.415	-1.867	-2.184
	Asymp. Sig.	.016	.062	.029

2 tailed Wilcoxon paired tests to compare the number of contexts produced in different writing tests in the samples where the Friedman test found statistical significance (or near)

All samples increased the number of contexts they produced between the pre and dp tests. The only sample to produce more contexts between post and dp tests was the merged PI group. Between the pre and post tests a few samples increased the number of contexts they produced: the whole data set, class C, the merged PI group and the PI group in class B. For class C this suggests that a test effect encouraged learners to produce more (there was no FonF/S between pre and post tests). For the Merged PI and PI in class B groups, it is possible that this indicates some tendency for PI to encourage class B learners to make more attempts to produce subject + verb structures. This was also implied to a more limited extent by the parametric tests above. This was not seen in class A or EI learners.

Appendix 36 Statistical procedures used to explore aspects of reliability and validity in the speaking tests

	Test	N	Mean	Std. Deviation
Class A	PRE	8	65.00	9.20
	POST	8	69.75	9.65
	DP	8	70.75	9.79
Class B	PRE	8	57.25	8.61
	POST	8	65.00	8.82
	DP	8	67.75	6.36
Class C	PRE	8	55.75	13.33
	POST	8	63.00	7.86
	DP	8	64.25	8.54
Overall	PRE	24	59.33	10.93
	POST	24	65.92	8.90
	DP	24	64.75	11.81

Mean obligatory contexts and standard deviations for oral narrative tasks, by class

The majority of the sample groupings can be considered to have normal distributions (using K-S tests and Shapiro-Wilks tests to determine the p value for some of the very small samples, see appendix 29), but as the n was very small in the smallest sample (e.g. PI in class A n=4), non-parametric tests were also carried out. Levene's test of equality of error variances showed that this assumption was upheld in the pre and post test data at the 95% confidence level and the dp test data at the 99% level.

	F	df1	df2	Sig.
PRE	2.165	4	19	.112
POST	1.017	4	19	.424
DP	3.304	4	19	.032

Levene's test of equality of error variances in the number of contexts produced in the oral pre, post and dp tests

Source	Type III Sum of Squares	df	Mean Square	F	Sig. ¹
CONTEXTS	606.787	2	303.393	10.440	.000
CONTEXTS * GROUP	81.167	2	40.583	1.396	.260
CONTEXTS * CLASS	46.167	2	23.083	.794	.459
CONT * GROUP * CLASS	10.167	2	5.083	.175	.840

Repeated measures ANOVA on the number of contexts produced in the oral tests

¹ Mauchly's Test of Sphericity showed that no adjustments were required for the calculation of the p value for the repeated measures ANOVA (Mauchly's W 0.805, approx Chi-square 3.904, 2, sig=0.142).

Q4 I think we have supposed to learn rules, which are shown on the OHP at the beginning of the lesson. I don't remember the rules though, I think it would have been better if we had made notes on the rules.....this helped us to learn vocabulary.

Processing Instruction learners

School A

I have learnt to look at the sentence more to see who exactly the sentence is talking about. I also find it easier now, after learning to look for odd letters at the end of verbs etc.

Yes it's made me look out for them and made me pay more attention to it

I remember things more if we do lots of activities on just one thing

They helped me to see the words that helped were a big part of the sentence that I didn't see before...they were different because they weren't just a lot of writing they made you pay attention to the important things.

I listen more for the endings and before I was using stuff like "le weekend dernier"

School B

I understand the French we learn better than before

It made you listen better as you had to listen to small endings and I know now how to right in the past tense better...I know what to look for at the ends of words

It is easier to pick out the little details

Q3 [yes they were different] focussed on specific parts whereas we used to just learn it more generally

I am more aware and notice things like the endings on verbs

They have taught me about what my mistakes were, and helped me to notice the difference (past + present)