AN ASSESSMENT OF EXPLICIT AND IMPLICIT INSTRUCTION IN THE CONTEXT OF THE INTERFACE DEBATE

by

Dimitrios Lysikatos

A thesis submitted in fulfilment for the Philosopher’s Degree in Theoretical and Applied Linguistics

ARISTOTLE UNIVERSITY OF THESSALONIKI
FACULTY OF PHILOSOPHY
SCHOOL OF ENGLISH LANGUAGE AND LITERATURE
DEPARTMENT OF THEORETICAL AND APPLIED LINGUISTICS
“Ο ενάρετος άνθρωπος δεν πρέπει να πουλά την ελεύθερη γνώμη του προσέχοντας πόσα χρήματα θα πάρει, και αν ακόμη είναι πολλά όσα του δίνουν. Γιατί τα πράγματα της ζωής αυτής είναι όμοια με όνειρο, και η φαντασία που προάγεται από τον πλούτο και άγνωστο είναι πού θα καταλήξει και λίγο καιρό μόνο κρατάει”.

(Άγιος Αντώνιος, ο Μέγας, Φιλοκαλία, Τόμος Α, σελ. 33, παρ. 37)
To the Silent Speakers
ACKNOWLEDGEMENTS

Carrying out a PhD Thesis requires spending valuable time in a solitary environment. My family have always been there offering their help in one way or another. Although many things have changed in my life during these laborious years, these people have been the ones that have kept me going throughout.

The present thesis would not have been possible without the guidance and assistance of Dr. Marina Mattheoudakis, my main supervisor. From the very beginnings of my initiation to the prospect of a PhD thesis, she has always been available offering valuable insights and recommendations throughout the project. I consider myself a very lucky person for having the opportunity to cooperate with Dr. Marina Mattheoudakis.

Dr. Eleni Agathopoulou and Dr. Despina Papadopoulou, the other two members of my supervisory committee, have contributed in ways that I could not sometimes understand, but now that the project is finished, I realize how important their contribution has been. I cannot overemphasize my gratitude to them.

Throughout, I have missed a lot, I have experienced a lot, and I have become wiser through recognizing my weak points and my ignorance. I deliberately made the decision to confine myself within the cloistered boundaries of my very close family members, my main supervisor, and the other two members of my supervisory committee. Thank you all!

I am grateful to God who answered my prayers even when I was too frustrated to hear the answers or too occupied to follow the path in front of me.
CONTENTS

Acknowledgments.................................................................iv
Contents.............................................................................v
Abstract.............................................................................xv
List of abbreviations............................................................xvii
List of tables......................................................................xix
List of figures.....................................................................xxii
List of histograms.............................................................xxiii
List of schemata...............................................................xxiv
List of diagrams...............................................................xxiv

CHAPTER 1: INTRODUCTION..............................................1

1.1. Explicit & Implicit Learning & Instruction......................1
1.2. Target Structure..........................................................3
1.3. Input Enhancement & Input Flooding............................4
1.4. Input Processing, Processing Instruction, & Structured Input 5
1.5. Research Questions......................................................6
1.6. Research Hypotheses..................................................6
1.7. Purpose....................................................................7
1.8. The Pilot Study & the Main Study..............................7
1.9. Aims and Significance of the Pilot Study & the Main Study 8
1.10. Thesis Outline.........................................................9
# CHAPTER 2: THE ENGLISH PAST TENSE

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>14</td>
</tr>
<tr>
<td>2.1. Tense &amp; Aspect</td>
<td>15</td>
</tr>
<tr>
<td>2.2. The English Past Tense</td>
<td>16</td>
</tr>
<tr>
<td>2.2.1. Past time reference in English</td>
<td>16</td>
</tr>
<tr>
<td>2.2.2. The English past tense: meaning, use &amp; form</td>
<td>16</td>
</tr>
<tr>
<td>2.3. The English Past Tense &amp; the Greek Past Perfective Form</td>
<td>18</td>
</tr>
<tr>
<td>2.3.1. The English past tense &amp; the Greek past perfective form: introduction</td>
<td>18</td>
</tr>
<tr>
<td>2.3.2. The English past tense &amp; the Greek past perfective form: differences &amp; similarities</td>
<td>19</td>
</tr>
<tr>
<td>2.3.3. The English past tense &amp; the Greek past perfective form: summary</td>
<td>22</td>
</tr>
<tr>
<td>2.4. The English Past Tense and the Greek Past Perfective Form: Some Remarks</td>
<td>22</td>
</tr>
<tr>
<td>2.5. The Acquisition of the English Past Tense</td>
<td>25</td>
</tr>
<tr>
<td>2.5.1. The acquisition of the English past tense: introduction</td>
<td>25</td>
</tr>
<tr>
<td>2.5.2. The acquisition of the English past tense: theory</td>
<td>27</td>
</tr>
<tr>
<td>2.5.3. The acquisition of the English past tense: research</td>
<td>27</td>
</tr>
<tr>
<td>2.5.4. The acquisition of the English past tense: summary</td>
<td>42</td>
</tr>
<tr>
<td>Conclusion</td>
<td>43</td>
</tr>
</tbody>
</table>

# CHAPTER 3: TEACHING GRAMMAR IN THE L2

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>44</td>
</tr>
<tr>
<td>3.1. Input Enhancement &amp; Input Flooding</td>
<td>45</td>
</tr>
<tr>
<td>3.1.1. Input enhancement &amp; input flooding: theoretical background: introduction</td>
<td>45</td>
</tr>
</tbody>
</table>
3.1.2. Input enhancement & input flooding: theoretical background........46
3.1.3. Input enhancement & input flooding: theoretical background; summary ...58

3.2. Input Processing & Processing Instruction........................................59
3.2.1. Input processing & processing instruction: theoretical background:
introduction.................................................................59
3.2.2. Input processing & processing instruction: theoretical background........60
3.2.3. Input processing & processing instruction: evolution......................71
3.2.3.1. Input processing & processing instruction: evolution: introduction.....71
3.2.3.2. Input processing & processing instruction: evolution: past & present....72
3.2.3.3. Input processing & processing instruction: evolution: research..........76
3.2.3.4. Input processing & processing instruction: evolution: final remarks.....77
3.2.4. Input processing & processing instruction: theoretical background:
summary.................................................................80

3.3. Input Enhancement & Input Flooding: Research.................................80
3.3.1. Input enhancement & input flooding: research: introduction.............80
3.3.2. Input enhancement & input flooding: research: studies.....................81
3.3.3. Input enhancement & input flooding: research: summary..................117

3.4. Input Enhancement, Input Flooding & the English Past Simple Tense....118
3.4.1. Input enhancement, input flooding & the English Past Simple Tense:
introduction.................................................................118
3.4.2. Input enhancement, input flooding & the English Past Simple Tense:
research.................................................................118
3.4.3. Input enhancement, input flooding & the English Past Simple Tense:
summary.................................................................125

3.5. Processing Instruction: Research..................................................126
3.5.1. Processing instruction: research: introduction..............................126
3.5.2. Processing instruction: research: studies.....................................128
3.5.3. Processing instruction: research: summary ........................................168

3.6. Processing Instruction & the English Past Simple Tense ..........................169

3.6.1. Processing Instruction & the English Past Simple Tense: introduction ......169

3.6.2. Processing Instruction & the English Past Simple Tense: research ..........169

3.6.3. Processing Instruction & the English Past Simple Tense: summary ..........171

3.7. Structured Input: Research ....................................................................171

3.7.1. Structured input: research: introduction ............................................171

3.7.2. Structured input: research: studies ....................................................172

3.7.3. Structured input: research: summary ................................................180

3.8. Structured Input & the English Past Simple Tense ..................................180


3.8.2. Structured input & the English Past Simple Tense: research ..............181

3.8.3. Structured input & the English Past Simple Tense: summary ..............183

Conclusion .....................................................................................................184

CHAPTER 4: METHOD ..................................................................................186

Introduction ..................................................................................................186

4.1. Research Questions and Hypotheses .....................................................187

4.2. Research Methodology ..........................................................................188

4.3. Instructor ..............................................................................................190

4.4. School Profile .......................................................................................191

4.5. Participants ............................................................................................192

4.6. Research Design ....................................................................................194

4.7. The Pilot Study ......................................................................................194
CHAPTER 5: RESULTS: INTERPRETATION

Introduction

5.1. Sample Analysis

5.2. Interpretation: Whole Sample

5.2.1. Interpretation: whole sample: mean performance & standard deviation values

5.2.2. Interpretation: whole sample: statistically significant differences

5.2.3. Interpretation: whole sample: summary

5.3. Interpretation: SI, IF, and C groups

5.3.1. Interpretation: SI, IF, and C groups: mean performance & standard deviation values

5.3.2. Interpretation: SI, IF, and C groups: statistically significant differences

5.3.3. Interpretation: SI, IF, and C groups: summary

5.4. Interpretation: Summary of Statistically Significant Differences

5.5. Interpretation: Individual Participants’ Performance

5.5.1. Interpretation: individual participants’ performance: introduction

5.5.2. Interpretation: individual participants’ performance: SI group

5.5.3. Interpretation: individual participants’ performance: IF group

5.5.4. Interpretation: individual participants’ performance: C group

5.5.5. Interpretation: individual participants’ performance: summary

5.6. Interpretation: Item Analysis

5.6.1. Interpretation: item analysis: introduction

5.6.2. Interpretation: item analysis: pre-test
6.5.2. Production: item analysis: pre-test ................................................................. 291
6.5.3. Production: item analysis: post-test ............................................................. 295
6.5.4. Production: item analysis: summary ............................................................ 298

6.6. Production: Error Analysis ........................................................................... 299

6.6.1. Production: error analysis: introduction .................................................... 299
6.6.2. Production pre-test & post-test: error analysis: whole sample ............... 300
6.6.3. Production pre-test & post-test: error analysis: whole sample: summary ..... 302
6.6.4. Production pre-test: error analysis: SI, IF, & C groups ....................... 302
6.6.5. Production pre-test: error analysis: SI group ............................................ 306
6.6.6. Production pre-test: error analysis: IF group ............................................ 306
6.6.7. Production pre-test: error analysis: C group ............................................ 307
6.6.9. Production post-test: error analysis: SI group .......................................... 308
6.6.10. Production post-test: error analysis: IF group ........................................ 309
6.6.11. Production post-test: error analysis: C group ........................................ 310
6.6.13. Production: error analysis: summary ...................................................... 311

Conclusion ............................................................................................................ 312

CHAPTER 7: DISCUSSION ...................................................................................... 314

Introduction ........................................................................................................... 314

7.1. Research Questions Revisited .................................................................... 315
7.2. Research Hypotheses Revisited .................................................................. 317
7.3. Discussion of Results: Interpretation .......................................................... 320
7.3.1. Discussion of results: interpretation: introduction ........................................321
7.3.2. Discussion of results: interpretation: tentative explanations .......................321
7.3.3. Discussion of results: interpretation: statistically significant differences ......324
7.3.4. Discussion of results: interpretation: general conclusions ............................325

7.4. Discussion of Results: Production .................................................................325
7.4.1. Discussion of results: production: introduction ...........................................325
7.4.2. Discussion of results: production: tentative explanations ..........................326
7.4.3. Discussion of results: production: statistically significant differences ..........327
7.4.4. Discussion of results: production: general conclusions ..............................327

7.5. Explanation of Results: Interpretation & Production: Summary ....................328

7.6. Discussion: The Control Group .................................................................329
7.6.1. Discussion: the control group: introduction ..................................................329
7.6.2. Discussion: the control group: explanations ..................................................331
7.6.3. Discussion: the control group: summary .....................................................342

7.7. Discussion: The Acquisition of the English Past Tense ..................................343
7.7.1. Discussion: The acquisition of the English past tense: introduction ..........343
7.7.2. Discussion: The acquisition of the English past tense: pertinent remarks ......344
7.7.3. Discussion: The acquisition of the English past tense: summary ..............353

7.8. Discussion: Teaching Grammar in the L2 ....................................................353
7.8.1. Discussion Teaching grammar in the L2: introduction ................................353
7.8.2. Teaching grammar in the L2: the main study in relation to other studies ......354

7.8.2.1. Discussion: Teaching grammar in the L2: the main study in relation to other studies: introduction .................................................................354

7.8.2.2. Discussion: Teaching grammar in the L2: the main study in relation to other studies: input enhancement & input flooding ........................................355

7.8.2.3. Discussion: Teaching grammar in the L2: the main study in relation to other studies: structured input & processing instruction .................................359
7.8.2.3. Discussion: Teaching grammar in the L2: the main study in relation to other studies: summary……………………………………………………………………364

7.8.3. Teaching grammar in the L2: summary……………………………………364

7.9. Discussion: Teaching the regular English Past Simple Tense -ed to Young Greek L1 EFL learners……………………………………………………………365

7.9.1. Discussion: Teaching the regular English Past Simple Tense -ed to Young Greek L1 EFL learners: introduction………………………………………………365

7.9.2. Discussion: Teaching the regular English Past Simple Tense -ed to Young Greek L1 EFL learners: curricular issues………………………………………………366

7.9.3. Discussion: Teaching the regular English Past Simple Tense -ed to Young Greek L1 EFL learners: summary………………………………………………….369

7.10. Discussion: Implications Language Teaching……………………………..370

7.11. Discussion: Implications for Teaching Young Learners…………………370

7.12. Discussion: Implications for Implicit Learning……………………………..371

7.13. Limitations………………………………………………………………………..372


GENERAL CONCLUSION…………………………………………………………………...376

BIBLIOGRAPHY………………………………………………………………………………377
ABSTRACT

The present study extends previous research on implicit learning and teaching conditions and compares the relative effects of 2 types of instruction, i.e. Structured Input (labelled SI), and Typographical Input Enhancement (TIE) & Input Flood (labelled IF), targeted to the acquisition of the regular English Past Simple Tense –ed by young Greek L1 3rd-grade Primary State School learners of English. A pilot study and a main study were designed and implemented. Only the findings of the main study are reported in the present thesis. Accordingly, in the main study 41 8-year-old classroom learners of English were assigned to 1 of 2 instructional treatment conditions: SI group, TIE & IF (IF group). A Control (C) group of learners who only participated in the assessment measures was also incorporated in the research design. Pre-tests and post-tests included interpretation and production tasks. These tasks were very similar to those used in research within the Input Processing (IP) and Processing Instruction (PI) paradigm commonly addressed to adult L2 learners. Consequently, a major objective of the present study was to document whether young L2 learners could successfully respond to the SI and IF & TIE instructional treatment conditions and the aforementioned assessment measures, i.e. whether such research designs can assist young learners acquire a grammatical structure beyond their current proficiency level. The data obtained were submitted to detailed statistical analyses (SPSS). These statistical analyses not only aimed at measuring learners’ performance per se, but also at validating the design and suitability of the assessment measures as to their efficacy in assessing these learners’ performance. An unexpected finding of the present study concerned the results obtained by the learners in the C group who exhibited enhanced learning performance as to both the interpretation and the production of the target structure. The overall results indicated that (a) the aforementioned treatment conditions can be successfully applied to teaching young Greek EFL learners, (b) young Greek EFL learners can learn grammatical forms beyond their proficiency level if such treatment conditions are adapted to their age and current proficiency level, (c) interpretation and production are 2 distinct processes at this age, and (d) such
learners profit the most from unobtrusive and highly implicit learning techniques. The study adds to the EFL literature focusing on learning at a young age.
LIST OF ABBREVIATIONS

C.....Control (group)
CPH.....Critical Period Hypothesis
CR.....Consciousness Raising
DG.....Dictogloss
EFL.....English as a Foreign Language
EI.....Explicit Information
EnI.....Enriched Input
EOI.....Extended Optional Infinitive (stage)
ESL.....English as a Second Language
FLA.....Foreign Language Acquisition
FMC.....Form Meaning Connection
FonF.....Focus on Form
Fonfs.....focus on forms
GI.....Grammar Interpretation
MLU.....Mean Length of Utterance
IE.....Input Enhancement
IF.....Input Flooding
IL.....Interlanguage
IP.....Input Processing
L1.....First Language
L2.....Second Language
LI.....Language Impairment
MOI, MOBI.....Meaning-Output-based Instruction
OBI.....Output-Based Instruction
OI.....Optional Infinitive (stage)
OP.....Output Processing (instruction)
PI.....Processing Instruction
PB.....Production Based Instruction
PIE.....Processing Instruction plus Enhancement
PIEcomp.....Processing Instruction plus Enhancement delivered through computers
PPP.....Presentation Practice Production
SI.....Structured Input
SIAE.....Structured Input Activities Enhanced
SIA.....Structured Input Activities
SLA.....Second Language Acquisition
SLI.....Specific Language Impairment
TD.....Typically Developing (children)
TE.....Textual Enhancement
TIE.....Textual Input Enhancement
VIE.....Visual Input Enhancement
WM.....Working Memory
ZPD.....Zone of Proximal Development
LIST OF TABLES

CHAPTER 4: METHOD

Table 1: Participant distributions.................................................................193
Table 2: Timing of main study instructional treatment tasks.........................225
Table 3: Duration of main study assessment measures.................................226

CHAPTER 5: RESULTS: INTERPRETATION

Table 4: Interpretation pre-test and post-test mean performance value and standard deviation value (whole sample)..........................................................231
Table 5: Interpretation pre-test and post-test normality tests (whole sample)...235
Table 6: Interpretation post-test to pre-test statistical analyses (whole sample)..................................................................................................................236
Table 7: SI, IF, and C group interpretation pre-test and post-test standard deviation values.................................................................239
Table 8: Interpretation: SI, IF, and C groups: Shapiro-Wilk normality tests...240
Table 9: Interpretation pre-test t-tests comparing the SI, IF, and C groups....242
Table 10: interpretation post-test test statistics comparing the SI, IF, and C groups..............................................................................................................244
Table 11: Interpretation SI group and IF group post-test to pre-test statistical analyses.................................................................247
Table 12: Interpretation C group pre-test to post-test statistical analyses......248
Table 13: Interpretation: summary of statistically significant differences......249
CHAPTER 5: RESULTS: PRODUCTION

Table 22: Production pre-test and post-test mean performance value and standard deviation value (whole sample).........................................................266

Table 23: Production pre-test and post-test normality tests (whole sample).....269

Table 24: Production post-test to pre-test statistical analyses (whole sample)...270

Table 25: SI, IF, and C group interpretation pre-test and post-test standard deviation values.................................................................273

Table 26: Production: SI, IF, and C groups: Shapiro-Wilk normality tests.....274

Table 27: Production pre-test test statistics comparing the SI and IF instructional treatment conditions.........................................................276

Table 28: Production pre-test t-tests comparing the SI and C groups.........276

Table 29: Production pre-test test statistics comparing the IF and C groups...276

Table 30: Production post-test t-tests statistics comparing the SI, IF, and C groups......................................................................................279

Table 31: Production SI group pre-test to post-test statistical analyses.......282
Table 32: Production IF group post-test to pre-test statistical analyses………283
Table 33: Production C group pre-test to post-test statistical analyses………284
Table 34: Production: summary of statistically significant differences………285
Table 35: SI group participants’ performance in the production tests………..286
Table 36: IF group participants’ performance in the production tests………..288
Table 37: C group participants’ performance in the production tests………..289
Table 38: Summary of SI, IF and C group participants’ performance in the production tests………………………………………………………………………………………………………………290
Table 39: Production pre-test item analysis……………………………………292
Table 40: Production pre-test item analysis: summary…………………………293
Table 41: Production post-test item analysis……………………………………295
Table 42: Production post-test item analysis: summary…………………………296
Table 43: Production pre-test error analysis (all groups)…………………………300
Table 44: Production post-test error analysis (all groups)…………………………300
Table 45: Production pre-test error analysis (each group separately)…………304
Table 46: Production post-test error analysis (each group separately)…………304
Table 47: Production: error sequential patterns: summary……………………311

CHAPTER 7: DISCUSSION

Table 48: Summary of explanations of C group performance…………………..343
LIST OF FIGURES

CHAPTER 3: TEACHING GRAMMAR IN THE L2

Figure 1: Stages in learning based on input..............................................46
Figure 2: Continuum of IE techniques from more explicit to less explicit......47
Figure 3: Processes in SLA........................................................................75
Figure 4: Traditional grammar instruction................................................75
Figure 5: Processing grammar instruction..................................................75

CHAPTER 4: METHOD

Figure 6: Pilot study research schedule......................................................204
Figure 7: Main study research schedule......................................................225

CHAPTER 5: RESULTS: INTERPRETATION

Figure 8: SI, IF, and C group interpretation pre-test and post-test mean performance values.................................................................238

CHAPTER 6: RESULTS: PRODUCTION

Figure 9: SI, IF, and C group production pre-test and post-test mean performance values...............................................................................272
Figure 10: Sequential pattern of types of errors in the production pre-test (all groups)...........................................................................301
Figure 11: Sequential pattern of types of errors in the production post-test (all groups)…………………………………………………………………………………………………………..302

Figure 12: Sequential pattern of types of errors in the production pre-test (SI group)…………………………………………………………………………………………………………….306

Figure 13: Sequential pattern of types of errors in the production pre-test (IF group)……………………………………………………………………………………………………..307

Figure 14: Sequential pattern of types of errors in the production pre-test (C group)……………………………………………………………………………………………………..307

Figure 15: Sequential pattern of types of errors in the production post-test (SI group)……………………………………………………………………………………………………..309

Figure 16: Sequential pattern of types of errors in the production post-test (IF group)……………………………………………………………………………………………………..309

Figure 17: Sequential pattern of types of errors in the production post-test (C group)……………………………………………………………………………………………………..310

LIST OF HISTOGRAMS

CHAPTER 5: RESULTS: INTERPRETATION

Histogram 1: Interpretation pre-test performance frequencies for the SI, IF, and C groups combined (normal distribution bell curve line included)………………232

Histogram 2: Interpretation post-test performance frequencies for the SI, IF, and C groups combined (normal distribution bell curve line included)………233

CHAPTER 5: RESULTS: PRODUCTION

Histogram 3: Production pre-test performance frequencies for the SI, IF, and C groups combined (normal distribution bell curve line included)………………267
Histogram 4: Production post-test performance frequencies for the SI, IF, and C groups combined (normal distribution bell curve line included) ................. 267

LIST OF SCHEMATA

CHAPTER 2: THE ENGLISH PAST TENSE

Schema 1: Time Line for the English Past Simple Tense .........................17

LIST OF DIAGRAMS

CHAPTER 7: DISCUSSION

Diagram 1: Likelihood of post-test performance of imaginary subject who scored 70% in the pre-test ..................................................................................340

Diagram 2: Imaginary normal distribution diagram including all subjects and using combined pre-test and post-test mean scores as reference points ...........341
CHAPTER 1:

INTRODUCTION

1.1. Explicit & Implicit Learning & Instruction

Pedagogy has always been interested in finding optimal ways of learning and teaching. One of the issues learning theory and research have focused on is explicit and implicit learning and instruction. Recently, the issue has received a lot of attention within SLA. Many attempts to provide operant definitions have been made. A possible interface or non-interface between the 2 has been proposed. Research has attempted to examine the effects of explicit and implicit instruction. The present thesis is placed within the framework of implicit instruction.

Definitions of explicit and implicit learning and instruction are difficult to articulate, but necessary. As to explicit instruction, it involves the presentation of information on rules in the input to the learners. In implicit instruction instructors do not present learners with information on rules in the input (R. Ellis, 1994, p. 642). As to explicit learning, learners process the input consciously attempting to discover the underlying patterns according to which rules and concepts are governed. In implicit learning, learners process the input without such a conscious intention (Hulstijn, 2005, p. 131).
The existence of a possible link between explicit and implicit learning has been labeled the *interface debate*. Relevant theory and research includes 3 points of view: (a) the **strong interface position**, (b) the **weak interface position**, and (c) the **non-interface position**. According to the strong interface position, declarative knowledge can be converted into procedural knowledge through practice (DeKeyser, 1998). According to the weak interface position, explicit knowledge can be transformed into implicit knowledge and the other way round (R. Ellis, 2004). According to the non-interface position, there is no direct transfer between explicit learning (learning) and implicit learning (acquisition) (Krashen, 1981).

Research findings on explicit and implicit learning, instruction, and knowledge are not conclusive, but a number of points merit presentation. First, as to explicit instruction, (a) it leads to substantial gains, (b) it is more effective than implicit instruction, and (c) it procures durable effects (Norris & Ortega, 2000). Second, as to implicit learning, (a) it “is a relatively robust system that survives psychological, psychiatric, and neuroanatomical injury”, (b) it “shows relatively little interindividual variability, and” (c) it “is relatively unaffected by ontogenetic factors” (Reber et al. 1999, p. 504, qtd. in Hulstijn, 2005, p. 131).

Within this context, the present thesis explores the effects of implicit instruction on young Greek L1 EFL learners based on the following reasoning. The 1st reason behind focusing on implicit instruction was based on the assumption that young EFL learners lack the meta-linguistic knowledge and skills to profit from explicit rule presentations. The 2nd reason was that the teaching methodologies utilized in the pertinent studies have been rarely investigated with this age group. Despite the fact that the findings reported in the present thesis are liable to a multitude of interpretations, a number of conclusions are validated by the data analyses.
1.2. Target Structure

The regular English Past Simple Tense –ed was the target structure investigated in the present thesis. The reason behind this choice was based on a number of interrelated assumptions. The 1st assumption was that there is no direct equivalence between the English past tense and the Greek past perfective form. In other words, the L2 structure the learners were required to learn presents a number of dissimilarities with their equivalent L1 structure. The 2nd assumption was based on previous research that the regular English Past Simple Tense –ed is a language form that presents problems to EFL learners. In fact, Benati proposed that it is especially problematic for Greek L1 EFL learners (Benati, 2005, p. 76). The 3rd assumption was that the acquisition of this target structure involves the Lexical Preference Principle as proposed by theorizing on Input Processing. This principle states that “learners prefer processing lexical items to grammatical items (e.g. morphology) for the same semantic information” (VanPatten, 2002, p. 758). According to this principle, in a sentence such as Yesterday I played football, EFL learners will rely predominantly on the time phrase yesterday than on the suffix -ed in order to locate the action of the verb play in past time. The 4th and final assumption was that the regular English Past Simple Tense –ed morpheme is acquired relatively late according to L2 orders of acquisition (e.g. Dulay & Burt, 1973, p. 273), since learners can rely on other cues to ascertain the notion of pastness. In other words, the morpheme seems redundant in the presence of time markers. The implicit instructional methodologies utilized to provide instruction on the target structure were a combination of typographical input enhancement (TIE) and input flooding (IF), and structured input (SI).
1.3. Input Enhancement & Input Flooding

Sharwood Smith (1993) employed the term *input enhancement* (IE) as a way of increasing input saliency. He adopted the difference between *comprehension* and *production*, as well as the constructivist notion that knowledge needs to be compartmentalized into smaller units. For this to happen, input needs to be *enhanced* in order to become more salient and more easily learned. Sharwood Smith distinguished between *positive input enhancement*, i.e. making “more salient certain correct forms in the input” and *negative input enhancement*, i.e. flagging “given forms as incorrect, thus signalling to the learner that they have violated the target norms” (p. 177). According to White (1998), *typographical enhancement* (TE) is a more explicit *input enhancement* technique than *input flooding*, but less explicit than the provision of rules. In actual research and teaching practice, *typographical input enhancement* techniques have taken various forms such as *bolding, underlining, italicizing*, the *use of capital letters* etc. in order to make target forms in the input more salient to learners. Through these implicit instructional techniques it has been suggested that learners will notice target forms in the input and acquisition will be aided. This is the basic underlying assumption behind typographical input enhancement as an instructional technique.

*Input flooding* is another form of input enhancement. Input flooding is an instructional methodology that is based on the assumption that input that includes as many instances of the target forms as possible is a necessary and sufficient condition for acquisition. In research practice it usually takes the form of including as many
tokens of the target structure as possible in the instructional materials. Input flooding has been extensively utilized in various overt or covert forms in research designs. In fact, Godfroid argued that even though “it is not usually labelled that way, input flood has been one of the preferred instructional techniques in laboratory studies on incidental or implicit learning” (Godfroid, 2015, p. 5).

1.4. Input Processing, Processing Instruction, & Structured Input

Structured input (SI) is a component of Processing Instruction (PI) which is the practical implementation of an extended input-based model of SLA labelled Input Processing (IP) (Lee & VanPatten, 1995; VanPatten, 1993, 1996, 2000, 2002; VanPatten & Cadierno, 1993). Processing Instruction is based on the tenet that second or foreign language learning is in some cases hindered by learners’ non-optimal learning strategies. These strategies need to be altered for acquisition to take place. In this respect, a number of input processing principles have been formulated (VanPatten, 2002, p. 758). According to theorising on Input Processing, learners have to be instructed in order to abandon these erroneous strategies through the provision of explicit information and / or structured input activities. This means that Input Processing is applicable only to the teaching of target forms that involve such principles and it cannot be applied to instruction of target forms that do not involve the alteration of such erroneous strategies. Processing Instruction attempts to practice the pre-selected forms by means of structured input activities. The 2 essential characteristics of such activities are that learners encounter forms while they are focused on meaning and that at no point do they practice producing forms; they only process them (Lee & VanPatten, 1995, p. 102). Lee and VanPatten divided structured
input activities in 2 major categories: (a) *referentially oriented*, i.e. *referential*, which “use an immediate concrete reference to ascertain the truth-value of a sentence” (e.g. students choose 1 of 2 sentences that best matches a given picture) and (b) *affectively oriented*, i.e. *affective*, “that ask for an opinion, require a personal response, or, in general, tap the student’s own world” (e.g. surveys) (p. 109).

1.5. Research Questions

The present thesis aimed at answering 2 main research questions:

The 1st research question inquired whether the SI and IF & TIE implicit methodologies are appropriate for the acquisition of the regular English Past Simple Tense *-ed* target structure.

The 2nd research question inquired whether the SI and IF & TIE implicit methodologies are appropriate for the acquisition of the target structure by young Greek L1 EFL learners.

1.6. Research Hypotheses

Two hypotheses underlying the main study were formulated:

The 1st hypothesis was that implicit instruction (in the form of the SI and IF & TIE instructional methodologies) can help young Greek L1 EFL learners acquire the regular English Past Simple Tense *-ed* structure.

The 2nd hypothesis was that young Greek L1 EFL learners learning the regular English Past Simple Tense *-ed* after being subjected to the 2 different implicit instructional treatments will benefit the most from the instructional treatment that is
the most explicit-like and both will outperform a 3rd comparison, control group. More specifically, young Greek L1 EFL learners learning the regular English Past Simple Tense -ed after being subjected to the more explicit-like implicit instructional treatment of structured input (SI Group) will benefit more than young Greek L1 EFL learners learning the regular English Past Simple Tense -ed after being subjected to a more implicit instructional treatment combining input flooding and textual input enhancement (IF Group). Both groups will outperform the 3rd group (C group, control) that will take part only in the assessment measures without any kind of exposure to the target structure.

1.7. Purpose

The purpose of carrying out the research to be presented was to examine the effectiveness of implicit instructional methodologies, i.e. SI and IF & TIE, as to the acquisition of the regular English Past Simple Tense –ed by young Greek EFL learners.

1.8. The Pilot Study & the Main Study

Based on the theoretical and research frameworks of input enhancement and input flooding, and input processing, processing instruction and structured input, a pilot study and a main study were designed and implemented. The research designs of both studies were almost identical.

In both studies 2 experimental groups of learners were formed. First, a group of learners labelled SI. These learners were exposed to structured input activities.
Second, a group of learners labelled IF. These learners were exposed to *input flooding* and *typographical enhancement* in the form of *bolding*. A third group of learners labelled C acted as a control / comparison group. These learners only took part in the assessment measures. These measures were interpretation and production pre-tests and post-tests.

**1.9. Aims and Significance of the Pilot Study & the Main Study**

The aims of both the pilot study and the main study affected both the choice of target structure and the participant age group. The target structure chosen was the regular English Past Simple Tense *-ed* and the participants were 3\textsuperscript{rd} grade Greek State Primary School pupils aged 8 years old.

The 1\textsuperscript{st} aim was to examine whether any subject (a grammatical structure in this case) can be taught effectively to all / any learner(s) of whatever age at whatever level of knowledge / acquisition if organized, presented, and taught in accordance with the mental and cognitive capacities of the learners. This requires that the material to be acquired be adapted to the needs, capacities, and intelligence(s) of the learners and not vice-versa.

The 2\textsuperscript{nd} aim was to examine the efficacy of various types of *implicit instruction*. Implicit instruction means that the teacher’s main preoccupation is not to explain the particularities of the subject matter, but to aid the learners discover on their own the principles of knowledge.

The significance of both the pilot study and the main study is that the aforementioned instructional methodologies (i.e. SI and IF & TIE) have been
relatively under-researched within the context of the aforementioned age group (i.e. young EFL learners) and target structure (i.e. regular English Past Simple Tense – *ed*).

### 1.10. Thesis Outline

Chapter 2 is titled *The English Past Tense*. It begins by examining how past time temporal reference is expressed in English, which was the participants’ L2, and in Greek, which was the participants’ L1. This examination involves how the English regular Past Simple Tense and the equivalent Greek Aorist Tense (past perfective form) are formed, what they mean, and how they are used. Finally, English past tense acquisition studies with an emphasis on young learners are reviewed. Chapter 2 includes 5 sections. In section 2.1. the terms *tense* and *aspect* are introduced. In section 2.2. the *meaning*, *use*, and *form* of the English past tense are presented. In section 2.3. the English past tense and the near equivalent Greek past perfective form are compared, and differences and similarities are outlined. In section 2.4. the relevance of the discussion in the previous sections (i.e. sections 2.1, 2.2., and 2.3.) to the studies in the present thesis is described, focusing on a number of pertinent methodological decisions incorporated in the studies reported in the present thesis. In section 2.5. studies focusing on the acquisition of the English past tense are presented, and an attempt is made to document the major problems learners face as to the acquisition of this grammatical form. Emphasis is placed on findings relevant to young learners.

Chapter 3 is titled *Teaching Grammar in the L2*. It reviews the pertinent literature, including both theory and research. The umbrella topics investigated are *input enhancement* (IE) and *input flooding* (IF), *input processing* (IP) and *processing*
instruction (PI) and structured input (SI). Chapter 3 includes 8 sections. In section 3.1. the theoretical foundations of input enhancement and input flooding are presented. In section 3.2. the theoretical foundations of input processing, processing instruction, and structured input are outlined. In section 3.3 research studies within the input enhancement and input flooding frameworks are reviewed. Section 3.4. includes a presentation of studies on input enhancement and input flooding using as a target structure the English Past Simple Tense. In section 3.5. research studies within the processing instruction framework are reviewed. Section 3.6. includes a presentation of studies on processing instruction using as a target structure the English Past Simple Tense. In section 3.7. studies within the structured input framework are reviewed. Section 3.8. includes a presentation of studies on structured input using as a target structure the English Past Simple Tense.

Chapter 4 is titled Method. It reports on the implementation of 2 full-scale studies, i.e. a pilot study and the main study, that were carried out to examine the effects on acquisition of instructional treatment methodologies and techniques, i.e. SI, and IF & TIE, by young Greek L1 EFL State Primary School learners. Chapter 4 includes 9 sections. In section 4.1. the research questions and hypotheses of the present thesis are repeated. In section 4.2. the teaching methodologies and techniques employed in both studies are briefly explicated. In section 4.3. the profile and role of the instructor / researcher in both studies is described. In section 4.4. the profile of the school where both studies were carried out is outlined. In section 4.5. the participants of both studies are presented. In section 4.6. the research design of both studies is reported. In section 4.7. the pilot study is described. This section includes the following sub-sections: (a) short introduction, (b) tokens, (c) materials and implementation, (d) research schedule, (e) scoring, and (f) short summary. Section
4.8. illustrates in detail the changes that were incorporated in the main study based on the experience drawn from the pilot study. Section 4.9. describes the main study and includes the same sub-sections as section 4.7.

Chapter 5 is titled *Results: Interpretation*. It includes the results of the statistical analyses of the performance of the participants in the main study in the interpretation pre-test and in the interpretation post-test. Chapter 5 includes 6 sections. In section 5.1. the whole sample is presented. In section 5.2. the mean performance values and the standard deviation values of the whole sample are included and statistically significant differences between the interpretation pre-test and the interpretation post-test data set (including all 41 participants) are reported. In section 5.3. the mean performance values and the standard deviation values of each of the 3 groups (SI, IF, and C) separately are included and statistically significant differences of the 3 groups separately are reported. In section 5.4. the statistically significant differences in all data sets of the interpretation tests are summarized. In section 5.5. individual participants’ performance is analyzed. In section 5.6. an item analysis of the tokens in the interpretation pre-test and in the interpretation post-test is included.

Chapter 6 is titled *Results: Production*. It includes the results of the statistical analyses as to the performance of the participants in the main study in the production pre-test and in the production post-test. Chapter 6 includes 6 sections. In section 6.1. the mean performance values and the standard deviation values of the whole sample are included and statistically significant differences between the production pre-test data set and the production post-test data set (including all 41 participants) are reported. In section 6.2. the mean performance values and the standard deviation values of each of the 3 groups (SI, IF, and C) separately are included and statistically significant differences as to the performance of the 3 groups separately are reported.
In section 6.3. the statistically significant differences in all data sets of the production tests are summarized. In section 6.4. individual participants’ performance is analyzed. In section 6.5. an item analysis of the tokens in the production pre-test and in the production post-test is included. In section 6.6. an analysis of the types of errors the participants made in the production pre-test and the production post-test is presented.

Chapter 7 is titled *Discussion*. It includes an attempt to explain the findings of the statistical analyses, make tentative assumptions, and reach conclusions. Chapter 7 includes 14 sections. In section 7.1. the research questions are revisited. In section 7.2. the research hypotheses are revisited. In section 7.3. explanations of the results as to the interpretation of the target structure are provided. In section 7.4. explanations of the results as to the production of the target structure are provided. In section 7.5. general remarks as to both the interpretation and the production of the target structure are highlighted. In section 7.6. an attempt to provide explanations as to the performance of the control group is made. In section 7.7. the findings of the main study are discussed within the framework of the studies related with the acquisition of the English past tense reviewed in chapter 2. In section 7.8. the findings of the main study are cross examined with a number of studies on TIE, IF, and SI reported in chapter 3. In section 7.9. implications as to teaching the regular English past tense to young Greek EFL learners validated by the findings are outlined. In section 7.10. implications for language teaching validated by the findings are outlined. In section 7.11. implications as to the teaching of young learners are discussed. In section 7.12. implications for implicit earning validated by our findings are proposed. In section 7.13. the limitations of the present thesis are presented. In section 7.14. avenues for further research are suggested.
Each chapter, section, and the sub-sections include a short introduction and a short summary. The thesis ends with a general conclusion. The final part contains the bibliography. The appendices including all the materials are available on separate CDs.
CHAPTER 2

THE ENGLISH PAST TENSE

Introduction

The present section has 2 aims. The first aim is to examine how past temporal reference is expressed in English, which was the participants’ L2 in the main study, and in Greek, which was the participants’ L1 in the main study included in the present thesis. In this respect, presentation will clarify the similarities and the differences in the way the past is conceptualized in the 2 languages. Discussion will focus on how the English regular Past Simple Tense (-ed) and the equivalent Greek Aorist Tense (past perfective form) are formed, what they mean, and how they are used, i.e. discussion will focus on meaning, form, and use. It will become evident that as to the meaning and use, the 2 tenses have some subtle and distinct properties. Also, their form is governed by a different set of rules. The second aim is to review a number of studies investigating the acquisition of the English past tense. Such studies have attempted to explain the way the English past tense is acquired. They are reviewed in order to provide a better understanding as to how this tense is acquired in order to explain the findings of the main study in a more optimal way.

Chapter 2 includes the following sections. In section 2.1., the terms tense and aspect are introduced. In section 2.2., the meaning, use, and form of the English past tense are presented. In section 2.3., the English past tense and the near equivalent
Greek past perfective form are compared, and differences and similarities are discussed. In section 2.4., the relevance of the discussion in the previous sections to the studies in the present thesis is described, focusing on a number of pertinent methodological decisions incorporated in the studies. Finally, in section 2.5., research focusing on the acquisition of the English past tense is presented, and an attempt is made to present some major issues as to the acquisition of this tense. Special emphasis is placed on findings relevant to children learners.

2.1. Tense & Aspect

To begin with, the terms tense and grammatical aspect need to be introduced and explained. Tense is understood as referring to the association between the actual time an event occurs and reference time. Aspect, on the other hand, relates, among other things, to the time span between the beginning and the end of an action. In the case of tense, the focal point is on locating an event in the past, the present, or the future. In the case of grammatical aspect, the focal point is on the way the interlocutor views a situation or an event (Bardovi-Harlig, 2000, p. 10) which might take the form of an ascertainment of duration (imperfective aspect) or not (perfective aspect). Despite the fact that the term grammatical aspect cannot be accurately captured through the existence or not of duration, this latter term is used here for practical reasons (see e.g. Bardovi-Harlig, 2000, p. 96). On the issue, Robison (1990) uses Klein’s (1986) definitions; In Robison’s words,

“The tense of a situation is the relation between its associated interval and some reference time, either the time of utterance or a topic time determined by the preceding event or by a time adverbial. The English past tense, for
example, marks a situation as before the time of utterance. Aspect is the relation between a situation and just its associated interval on the time line. “Perfective aspect”, for example implies an exact correspondence between an event and its associated time interval, while “imperfective aspect” suggests a containment of the interval within the event.” (Robison, 1990, p. 317)

2.2. The English Past Tense

2.2.1. Past time reference in English

Temporal expression of the past in English is captured through both verbal morphology, as well as the use of adverbials. In the former case, according to Leech (1971, qtd in Bardovi-Harlig, 2000, p. 17), the past is manifested morphologically through the following Tenses, according to Example 1:

Example 1:

1. Past Simple Tense, e.g. I studied,
2. Past Progressive Tense, e.g. I was studying,
3. Present Perfect Simple Tense, e.g. I have studied,
4. Present Perfect Progressive Tense, e.g. I have been studying,
5. Past Perfect Simple Tense, e.g. I had studied,
6. Past Perfect Progressive Tense, e.g. I had been studying.

2.2.2. The English past tense: meaning, use & form

Bardovi-Harlig (2000, p. 178 - 180) used Reichenbach’s (1947) tense-aspect classification in order to explain the meaning of these different manifestations of past
time reference. In this respect, Reichenbach had discriminated among event time (E), speech time (S), and reference time (R) capturing the meaning of each Tense by plotting these 3 reference points on a time line. In the case of the English Past Simple Tense, the schematic time line could be roughly depicted as follows:

**Schema 1: Time Line for the English Past Simple Tense**

```
__________ Ð ______________
```

In the case of the English Past Simple Tense, event time (E) and reference time (R) coincide located in the past, whereas speech time (S) is located in the present. In other words, an imaginary speaker is talking now, at present (speech time), about an action which happened in the past (event time), referring to it (reference time) as belonging to the past, in the sense that it is completed and there is no kind of reference to the present speech time.

This interpretation means that the prototypical use of the English Past Simple Tense is to refer to actions that occurred at a specified time in the past and have been completed without any kind of explicit semantic reference to present time, i.e. even though the action might have present time consequences, the focal point is on the pastness of the action, its assumed completion (definiteness) in the past, and not its duration. More specifically, the English Past Simple Tense can be captured as including the following properties: [+ past], [+ realis], [+ perfective].
As to the formation of the regular English Past Simple Tense, it is morphologically expressed by the addition of the suffix \textit{–ed} to the base stem form of the infinitive, in its various allomorphs, e.g. /t/ as in \textit{talked}, /d/ as in \textit{listened}, and /-id/ as in \textit{started}. There is no morphological manifestation of either person, or number, or gender, or case. Person, number and (sometimes) gender are expressed through an obligatorily preceding subject pronoun, i.e. I, You, He/She/It, We, You, They, or an animate or inanimate subject, i.e. Dimitris, or Helen, or the dog. More specifically, the sole morphological manifestation of the regular English Past Simple Tense is the presence of the suffix \textit{–ed} to the base stem form which might be the sole indicator of [+ past], [+ realis], [+ perfective] in the absence of temporal expressions, or other linguistic and paralinguistic temporal markers like the chronological order of events. If such temporal signs are absent, the \textit{-ed} suffix becomes highly non-redundant and salient. In the case of the irregular English Past Simple Tense, the past tense form is realized through a number of irregular forms not always directly linked with the base verb form.

\textbf{2.3. The English Past Tense \& the Greek Past Perfective Form}

\textbf{2.3.1. The English past tense \& the Greek past perfective form: introduction}

The Greek equivalent to the English Past Simple Tense is the Αόριστος (Aorist) Tense, or, to be more accurate, the Greek past perfective form. The very term \emph{aorist} refers, if literally translated into English, roughly to something \textit{vague}, \textit{unspecific}, or, more accurately, \textit{indefinite}, in the sense that it cannot be exactly defined. In fact, the very terms \textit{past simple} and \textit{aorist} clearly signify that the
aforementioned notion of tense captures the prototypical sense of the English Past Simple Tense as referring to something that happened in the past, whereas the Greek Aorist operates more like an aspectual category since the focal point is on the way the interlocutor views a past situation or an event, as Bardovi-Harlig had commented previously (Bardovi-Harlig, 2000, p. 10). The ascertainment of duration (imperfective aspect) or not (perfective aspect) is, in this sense, irrelevant; the interlocutor simply refers to an event in the past in a non-definite way (αόριστα, a word which in English means roughly without specification, indefinitely).

2.3.2. The English past tense & the Greek past perfective form: differences & similarities

It has been suggested that the Greek Aorist can be translated into English through a variety of tenses, such as both the Past Simple Tense and the Present Perfect Tense, as the following example shows:

Example 2:

1. Μελέτησα τα μαθήματά μου.
   Studied-I my lessons.
2. I studied my homework.
3. I have studied my homework.

The issue of the correspondence between the 2 tenses has been of such consistent concern to translators that Robertson, referring to the Ancient Greek Aorist, has
commented that “(t)he aorist in Greek is so rich in meaning that the English labours and groans to express it. As a matter of fact the Greek aorist is translatable into almost every English tense except the imperfect” (Robertson, 1919, p. 847 – 848, parenthesis added). It is probably this indefiniteness of the Greek Aorist Tense that renders an exact one-to-one equivalence with the English Past Simple Tense impossible.

It is a fact that the uniqueness of the Greek Aorist has led to a multitude of tentative explanations related to its tense and aspectual properties. Most of the analyses regard the Greek verb tenses as including subtle interrelationships of both time and aspect with the latter being the predominant element (Wallace, 1996, p. 496). However, in the present thesis Lyons’ (1968, p. 314 -315) tripartite categorization of the Greek aspectual system will be embraced. This decision is arbitrary. It is, however, partly justified by the attempt to explain the properties of the Greek Aorist Tense in cross-comparison with the English Past Simple Tense using the terminology utilized to describe the latter, despite the fact that the ensuing discussion and generalizations are to be studied with great caution.

In this respect, Lyons proposed that the Greek aspectual system includes 3 instances based on the binary distinction between the perfective and imperfective aspect, and the durative and non-durative aspect. Accordingly, the perfective aspect encompasses the notions [+completive and - durative], the imperfective aspect encompasses the notions [-completive and + durative], and the aorist aspect encompasses the notions [-completive and - durative].

Lyons’ categorization captures an essential difference between the prototypical meanings / senses of the English Past Simple Tense and the Modern Greek Aorist Tense, i.e. that whereas in the former actions are conceptualized as complete, in the latter this element of completeness is not prominent. However, this
categorization is not language specific, which means that it is not an account based explicitly on the Greek language per se, but borrows its underlying principles from accounts drawn from other languages. For this reason, it is incomplete.

A more accurate language specific account was provided by Porter (1989) who proposed an explanation based exclusively on the Greek aspectual system in which tense is not considered at all. According to this account, the Greek verb system is characterized by 2 aspects, i.e. the aorist aspect and the non-aorist aspect, and the non-aorist aspect is further subdivided into the present non-aorist aspect and the perfect non-aorist aspect. Porter’s interpretation was based on the Ancient Greek language used in the Apocalypse of the Holy Bible.

Despite the fact that it might be criticized for disregarding tense, it is language specific, and, coupled with Lyon’s aforementioned account, captures the fundamental dissimilarities between the way the English language and the Greek language verbal past time referencing systems operate. If Porter’s and Lyons’ accounts are accurate, this might validly lead to the conclusion that Greek L1 EFL have a lot more to do than L1 to L2 transfer. In fact, in order to comprehensively acquire the English Past Simple Tense, they need to unlearn the Greek Aorist aspectual conceptualization and integrate into the English past time temporal tense system. It is not simply an issue of learning a novel item; it implies a firm decision to learn how to think in a different, foreign way.

As to the form of the Greek Aorist, to begin with, Greek is an inflectionally rich language. Morphologically, verbs in the Greek Aorist Tense are inflectionally marked for person, number, tense, voice, and mood. This is realized through combinations of the verbal stem with prefixes and suffixes. Unlike the English Past Simple Tense, an explicit overt subject is not obligatory as this information is
provided by suffixes on the verb stem. This represents a persistent problem for Greek L1 EFL learners who seem to insist on omitting overt subjects which, in most cases, are obligatory in English. To the contrary, these learners learn quite easily that the addition of the -ed suffix, necessary to form the English Past Simple Tense, even though they fail to produce it in spontaneous oral speech, up to the intermediate level.

2.3.3. The English past tense & the Greek past perfective form: summary

Combining such observations concerning the dissimilarities between the Greek Aorist Tense and the English Past Simple Tense as to their meaning and use coupled with their morphological dissimilarities already mentioned easily leads to the conclusion that Greek L1 EFL learners are justified in having problems acquiring the English Past Simple Tense.

2.4. The English Past Tense and the Greek Past Perfective Form: Some Remarks

The difficulties Greek EFL learners face when confronted with the English Past Simple Tense are further accentuated taking into account a fundamental dissimilarity between the Greek and the English language. Greek is an orthographically highly transparent language since the strings of phonemes that constitute a given word correspond to strings of letters or groups of letters in the same order very consistently (95% of cases for reading, 80% of cases for spelling) (Protopapas, 2014, p. 4). In contrast, English has a very deep (i.e. opaque) orthographic system making the relationship between letters and phonemes very arbitrary; in fact, this relationship is irregular in English. According to the
Orthographic Depth Hypothesis (see Katz and Frost, 1992, p. 150 – 151), shallow orthographies are more easily able to support a word recognition process that involves the language phonology. In contrast, deep orthographies encourage a reader to process printed words by referring to their morphology via the printed word’s visual-orthographic structure. In this respect, it has been proposed that the English language to some extent resembles the Chinese language as requiring the reader to think in visual idea units (see e.g. McBride-Chang et al., 2008, p. 172 for a review of studies that have reported commonalities between the English and the Chinese language). For languages with relatively deep orthographies, such as English, beginner readers have much more difficulty learning to decode words. As a result, learners of “orthographically opaque” languages learn to read more slowly than L1 learners of orthographically transparent languages like Greek (see Ellis, N. et al., 2004, p. 441). According to Ellis, N. et al. (2004), “transparent orthographies promote faster rates of reading acquisition” (p. 441).

More to the point, L1 Greek learners studying English as an L2, when confronted with the task of acquiring the English Past Simple tense face the additional challenge of figuring out the various allomorphs of the -ed suffix in their phonetic realizations, i.e. /t/, /d/, and /-id/, something which is not the case in the majority of cases in the orthographic system of Greek.

The aforementioned fundamental dissimilarities of the Greek Aorist Tense and the English Past Simple Tense as to their semantics, their use, and their formation, the striking dissimilarity as to the orthographic depth between the Greek language and the English language, and the young age and low proficiency of our learners, led to a number of important decisions as to the implementation of the research design of the main study.
First, the decision was made to use age-appropriate narratives including strict chronological sequencing of actions (verbs), since “chronological order is not restricted to learner language” (Bardovi-Harlig, 2000, p. 35), in an attempt to create a learning environment which would assimilate as closely as possible the commonalities in meaning and use of the two tenses; an environment where the meaning and use of the English Past Simple Tense would resemble as closely as possible relevant meanings and uses of the Greek Aorist Tense. The second decision was to eliminate any temporal expressions or other time indicators in order to render the tokens, and, more specifically, the –ed suffix, in the English Past Simple non-redundant and increase their salience. The third decision was to use visual typographical input enhancement to additionally increase the salience and non-redundancy of the target form. This would assumedly draw learners’ attention to the target form in an attempt to aid noting and processing, and subsequent acquisition since the various allomorphs may have posed problems for L2 learners. The fourth decision was to combine reading and listening in all parts of the research design, i.e. assessment tasks and treatment conditions, in order to obviate the differences in orthographic depth between the two languages and raise learners’ awareness as to the fact that the English Past Simple Tense -ed suffix has different phonetic realizations, a fact very uncommon in their L1. This final decision was not a pre-planned decision, but, rather, an on-the-spot decision, made by the researcher, instructor, and author of the present thesis during the implementation of the research design.

All these decisions were made taking into account the young age as well as the low level of proficiency in English of the participants in the study. In this respect, it was not attempted to force learners unlearn their L1 Aorist Tense conceptualization and provide explicit instruction on the English Past Simple Tense. Instead, it was
decided to create a learning environment that would resemble the meaning and use of the Greek Aorist Tense and, at the same time, implicitly introduce and aid learners induce the English Past Simple Tense in a naturalistic-like way.

Another issue that merits further discussion is that it was initially assumed that Greek learners would exhibit few problems in understanding - interpreting the English regular Past Simple Tense -ed. They would, however, have problems in forming - producing this tense since formation and production are quite different between the English regular Past Simple Tense -ed and Modern Greek Aorist Tense. This assumption was not substantiated by the results of the main study. In fact, a diametrically opposite state of affairs was borne out by the data, i.e. that accurately understanding - interpreting the English regular Past Simple Tense -ed was a burdensome task for Greek L1 learners. To the contrary, forming and producing the English regular Past Simple Tense by adding the novel -ed suffix to familiar verb stems was not as problematic for these learners (see chapters 5 and 6 for the results as to interpretation and production respectively).

2.5. The Acquisition of the English Past Tense

2.5.1. The acquisition of the English past tense: introduction

The English past tense has received considerable attention as the topic of investigation in recent years. The bulk of the research carried out has focused on the acquisition and processing by children and adults of its 2 distinct formation patterns, i.e. the productive regular –ed suffixation and the idiosyncratic irregular pattern forms. More specifically, the past tense of the majority of English verbs is formed by
adding the –ed suffix to the stem (e.g. play – played) and, according to a limited set of phonological rules, the past form verbs end in the allomorphs /t/, /d/, or /id/. This productive formation can be applied to thousands of English verbs. To the contrary, a limited set of around 150 to 180 verbs form the past tense in various non-systematic ways giving rise to irregular past tense forms such as go – went, hit – hit, burn – burnt, etc.

This apparent disparity between the so-called regular and irregular English past tense forms has attracted the attention of linguists, psychologists, psycholinguists, neurolinguists, etc. Such an enhanced attention is due to the belief that the understanding of the acquisition and processing of the regular and irregular past tense forms might offer insights as to how languages are acquired in general. The fact that past tense formation in English is inflectionally poor in relation to other languages has rendered this target form a case for the formation and testing of theoretical, research, and practical proposals. Such proposals, along with the most pertinent issues addressed are reviewed in the following sections. The importance of this review is that the research design and the findings of the main study reported in the present thesis, as well as the findings of relevant studies reviewed within the frameworks of input processing, processing instruction, structured input, typographical input enhancement, and input flooding in the field of SLA have rarely addressed the overarching issue of how regular and irregular past tense verbs are acquired and processed.

Even though English past tense irregular verbs were not investigated in the main study, reference will be made to research pertaining to this issue since the study of irregular past tense verb morphology is, most of the times, closely linked with that of regular past tense morphology. The explanations offered as to the acquisition and
processing of regular past tense verbs might offer valuable insights as to the discussion of the findings of the main study reported in the present thesis.

**2.5.2. The acquisition of the English past tense: theory**

Three theoretical models have addressed how morphological rules are stored and processed, also trying to account for the acquisition of the English past tense: (a) the dual-route processing model, (b) the connectionist model, and (c) the network model (Ambridge, 2010, p. 1497, Bybee, 1995, p. 425 – 428). First, according to dual-route processing models, English regular past tense forms are derived through a symbolic rule applied, i.e. adding the \(-ed\) suffix to the base stem form. On the other hand, English irregular past tense forms are lexical in nature requiring the search and finding of the appropriate lexical form. Second, according to connectionist models, no symbolic rules are required to from the regular past tense, but rather, all inflections, both regular and irregular, are formed in the same way through a network that maps base to past tense forms. Third, according to network models, the English past simple formation operates much like a pattern based on morphological associations the determinants of which are the type frequency and the token frequency of the actual verb forms.

**2.5.3. The acquisition of the English past tense: research**

Kuczaj (1977) investigated the extent of correct, incorrect, or nonuse of the English past tense inflection including both regular and irregular forms by young English speaking children aged between 2.6 and 5.6 years old during spontaneous
speech. Almost all the participants exhibited more correct use of the regular past tense form rather than the irregular past tense form, a finding which led Kuczaj to suggest that the regular form is more easily acquired than the irregular form, i.e. the regular form is acquired earlier (p. 593). Kuczaj reported that the types of overgeneralization errors, i.e. the tendency to add the –ed suffix to irregular verbs, varied with age. Accordingly, younger children tend to add –ed to the base form, i.e. eated*, whereas older children seem more prone to add –ed to the past tense form resulting in errors such as ated*.

Age is not the best predictor for correct production of regular past tense forms; for such forms mean length of utterance (MLU) is a more valid indicator, according to Kuczaj (p. 597). Kuczaj reasoned that this might be due to the fact that the regular past tense is rule-governed. He suggested that

“the reason that MLU is a better indicator of success on the regular past than is chronological age lies in the fact that some children apparently infer rules earlier (and perhaps on much more limited data) than do other children.” (Kuczaj, 1977, p. 598, parenthesis in the original)

MLU is not necessarily linked with rule-inference strategies, general learning abilities, or language acquisition. Kuczaj refers to this issue to explain why irregular past tense forms are acquired later and abstains from making such generalizations. MLU is often included in similar research designs. In my opinion, MLU is more of a character trait than a linguistic indicator. Some people are more talkative and eloquent than others. This does not mean that smaller MLUs are necessarily linked with lower acquisition patterns, despite Kuczaj’s findings. Language acquisition processes cannot be adequately encompassed by such a concept. Even though chronological age does not accord always accord with acquisition patterns, it is a more valid research tool in the sense that it can more easily be utilized as a standard to make comparisons.
Bybee and Slobin (1982) investigated patterns in the errors participants of 3 different age groups, i.e. preschoolers, 8-10 year old children, and adults, make as to the acquisition of English regular and irregular past tense forms. According to Bybee and Slobin children and adults make errors that exhibit similarities and differences (p. 267). This state of affairs led the 2 authors to suggest that both age groups use different schemas, i.e. generalizations based on phonological properties (p. 265). Bybee and Slobin based the discussion of their findings based on Zager’s (1980) distinction between source-oriented and product-oriented morphological modifications as to the creation of new words (Zager, 1980, qtd. in Bybee and Slobin, 1982, p. 285). Source-oriented modifications entail modifications of base forms in order to create derived forms and can account for how regular English past tense forms are derived. Product-oriented modifications lead to the production of words that resemble other words that belong to a similar morphological class. According to Bybee and Slobin, “English irregular past-tense forms make up product-oriented classes” described as schemas (p. 285).

Three different models that account for storing and processing of morphology were compared by Bybee (1995): the dual-processing model, the connectionist model, and the network model. Discussion at present will examine only the 3rd model as this is the one Bybee adhered to. The relationship between type frequency and token frequency are central in Bybee’s paper. According to Bybee’s network model, the underlying foundation of morphological properties and patterns are “associations made among related words in lexical representation” (p. 428). In other words, this model tries to account for morphological storage and processing on the basis of interactions among words in the lexicon. Such relationships emerge as a result of token frequency, i.e. how frequent a word is in the input, as well as type frequency, i.e.
connections of a word with other words on the basis of phonological, semantic, and morphological similarities and differences (p. 428 – 429).

The errors EFL learners or native English children make in the comprehension and production of the past tense may offer more valuable insights than their correct performance, as to the acquisition of this target form. Any description of the acquisition of the English past tense needs to account for such errors. For example, the provision of bare stem forms, instead of –ed inflected forms, does not necessarily implicate lack of knowledge of finiteness, according to Rice and Wexler (1996). Despite the fact that Rice and Wexler were more interested in discovering grammatical markers for both scientific and clinical reasons, their Extended Optional Infinitive (EOI) Stage, based on Wexler’s (1994, qtd. in Rice & Wexler, 1996, p. 1241) Optional Infinitive (OI) Stage (Rice & Wexler, 1996, p. 1241, 1243) may prove informative for our present purposes.

More specifically, Rice and Wexler summarized that their research findings as to the –ed suffixation process by 3- to 5-year old children with Specific Language Impairment (SLI) and typically developing (TD) children revealed that the inability of children with SLI to accurately produce finite past tense forms is characterized by oscillation between the suppliance of bare stem forms and the suppliance of the appropriate suffixed forms (p. 1247). Consequently, they abstained from unequivocally concluding on tense marking as a clinical marker (p. 1254) even though their research generally points to this direction.

However, in similar study, Rice, Wexler, and Cleave (1995) had already reported that –ed suffixation data “unequivocally” differentiated children with SLI from typically developing age- and mean-length-of-utterance (MLU) peers (Rice et al., 1995, p. 856). To be more precise, Rice et al. had summarized their findings as
indicating a “preference” for children with SLI for suppletion of bare stem forms, a preference not necessarily implying that such a choice is conscious (p. 862).

Jaeger, Lockwood, Kemmerer, Van Valin, Murphy, and Khalak (1996) attempted to address the issue whether English regular past tense verb forms and English irregular past tense verb forms are acquired through the same processes or not. Working within a neurolinguistic perspective, their method of investigation was positron emission tomography. The authors tried to document the brain activation patterns during the production of regular, irregular, and nonce verbs.

Superficially, the acquisition of the English past tense is a straightforward grammatical phenomenon including a general rule for regular verb forms, i.e. the addition of the –ed suffix, and a limited number of irregular verb forms to be learned ‘by heart’. On this basis, the English past tense has received undue attention by researchers from various related disciplines linked to the psychology of learning. However, there is

“much more at stake than the correct analysis of a relatively simple aspect of the grammar of a morphologically impoverished language. Indeed, the ultimate outcome of the debate has far-reaching implications for such major issues as cognitive architecture, innateness, and the psychological reality of linguistic rules and structures.” (Jaeger et al., 1996, p. 453)

If this account is correct, it means that studies investigating the acquisition of the English past tense tap on more fundamental issues than simply discovering the best way to ‘teach’ a target from. Finding the optimal way to help learners acquire the English past tense might be a piecemeal part of the larger picture of the attempt to explain learning and the human mind itself. The argument also works the other way round. Learning theories and explanations of how the human mind works might help shape and design appropriate teaching methodologies. In reality, then, the acquisition
of the English past tense is a complicated issue involving much more than simply using the correct regular or irregular past tense form to ‘discuss what happened’.

Jaeger et al.’s study could be summarized along the following lines. The participants were required to see 5 lists of words and provide oral responses. One of the lists included base stem forms of regular past tense verbs and the participants were asked to orally produce the respective past tense forms. Another list included base stem forms of irregular past tense verbs and the participants were also asked to orally produce the respective past tense forms. Data collection included brain activation areas and levels, oral responses, as well as reaction times (RTs). Each list included 46 verbs or nonce forms.

The list with the regular past tense verbs included a number of verbs which were also included as target tokens in the main study reported in the present thesis. These are the following 7 verbs in order of presentation in Jaeger et al.’s experiment: love, visit, jump, rub, look, watch, and walk (p. 490 – 491). This means that Jaeger et al.’s findings reported in the following paragraph may very tentatively underscore some details of the findings and their explanations of the present main study.

The most important findings of Jaeger et al. were as follows. First, response times (RTs) for regular past tense verbs were significantly smaller than RTs for irregular past tense verbs, and irregular nonce verbs (p. 465). Second, the task including regular past tense forms consumed the smaller amount of brain activation compared to the other conditions (p. 468 – 469). Third, the task including regular past tense forms involved the smaller brain area compared to the other conditions (p. 470). Overall, different brain areas were involved during the tasks including regular and irregular verbs.
Within a different research perspective, *productivity* in the case of the English past tense refers among other things to the provision of the regular –*ed* suffix to irregular base verb forms resulting in forms like *putted*, according to Marchman (1997, p. 283). This can also be linked to regular past tense inflection in an indirect way. The link is that Marchman argued that such productive irregular form errors are due to combinations of item frequencies, phonological properties of base and past tense forms, as well as what the author terms as phonological past tense “neighbourhoods” (op. cit.). Although Marchman utilized such notions to explain productivity of irregular past tense morphology by school-aged children, links with regular past tense morphology might be inferred and Marchman did use a small number of regular past tense verbs in her study.

*Neighborhood* in this context was “defined with respect to phonological attributes that are proposed to be relevant to the clustering of past tense forms” (Pinker & Prince, 1988, qtd in Marchman, 1997, p. 286). In other words, this concept is used to describe similar phonological properties of words. Within this context, past tense forms might have *friends*, i.e. other forms with similar base forms to past tense formation patterns, and *enemies*, i.e. other base forms with which they share phonological properties but their past tense forms differ (p. 286 – 287). The whole issue is based on the assumption that such *friends* and *enemies* either create groups among past tense forms or separate specific past tense forms from others thereby affecting acquisition.

Marslen-Wilson and Tyler (1998) made the point that neuropsychological data suggest that the 2 types of forms, i.e. regular and irregular, may share more similarities than differences. They argued
“that the underlying relationship between an irregular past tense and its stem is more like the morphological relationship between a regular inflected form and its stem than it is like the relationship between pairs that are just semantically related.” (Marslen-Wilson & Tyler, 1998, p. 433)

It needs to be emphasized that for our findings on regular past tense forms to be generalizable to irregular past tense forms, a partial replication study including irregular tokens is necessary. The argumentation put forth by Marslen-Wilson and Tyler may offer tentative credence to attempts to theoretically provide a unified account of both variants of the English past tense, i.e. regular and irregular. Nonetheless, the actual neuropsychological data of Marslen-Wilson and Tyler revealed that different brain areas were involved during the processing and acquisition of regular and irregular past tense morphology by patients with neuropathological deficiencies (p. 432).

Inflections including regular past tense forms were also included in Alegre and Gordon’s (1999) 1st experiment. Definitive conclusions as to the regular English past tense form on the basis of this study cannot be drawn since the target items included other grammatical forms as well. However, some of their findings might be relevant for our present purposes. The first was that RTs in their 1st experiment were affected by word length, i.e. the lengthier the word, the more time it took participants to respond (p. 46 – 47). If higher RTs are to be tentatively equated with enhanced difficulty in processing and production, this could lead to the hypothesis that the young children in our study should be more successful with regular past tense forms including few letters than regular past tense forms including a larger number of letters. Alegre and Gordon’s 2nd finding – pertaining to a host of regular inflections including regular past tense forms was that word frequency effects were operant only above the threshold level of approximately 6 per million words (p. 56 – 57). This led
the 2 authors to assume that regularly inflected words with frequencies above this threshold level may be retrieved as whole words, rather than generated through the application of rules. In light of our study, this might be interpreted as rendering only very highly frequent regular past tense forms easier to interpret and produce than forms of lower frequency. Medium to low frequency forms in our study should require the application of the –ed suffixation rule rendering them more difficult to interpret and produce.

Miozzo’s (2002) neuropsychological study of an anomic patient as to her processing of English regular and irregular past tense forms is not very easily comparable to our study. However, 2 points referred to by Miozzo merit discussion. The first is related with the anomic patient’s errors with regular past tense forms, whereas the second concerns the issue of imageability. Both issues will be reiterated in section 7.6.

According to the Aspect Hypothesis (Andresen & Shirai, 1994, qtd. in Collins, 2007, p. 296 – 297) L2 learners acquire the tense-aspect system of the target language on the basis of associations between tense-aspect forms and categories of verbs which best match such forms. This account proposes 4 (sometimes overlapping) semantic categories of verbs: (a) states, (b) activities, (c) accomplishments, and (d) achievements. Collins posited and provided experimental proof that English simple past tense forms are more frequently and accurately used with verbs belonging to the accomplishment and achievement categories by learners from different L1 backgrounds (p. 297 – 298). In fact, both L1 Japanese and L1 French participants of her study exhibited a similar tendency to supply tenses other than the simple past with verbs belonging to the state and activity verb classes in contexts that required the provision of the simple past (p. 299).
The regular English simple past tense was not directly investigated as a target form by Shak and Gardner (2008) who were interested about the attitudes of young learners on 4 different tasks: consciousness-raising (CR), dictogloss (DG), grammar interpretation (GI), and grammaring. However, all 3 tasks included the English past tense in one way or another. Consequently, the findings may be indicative as to the way learners prefer to be taught this target form. Positive attitudes and preferences enhance learning outcomes, in most cases.

Shak and Gardner’s grammar interpretation (GI) task targeting (past) active and passive forms shared a number of common elements with the interpretation tests utilized in our study the most notable of which was that learners were required to select sentences that matched pictures (p. 390, see also p. 406 – 407 in the appendix section). Learner preferences were assessed on 2 different occasions. The tasks were not identical on Day 1 and Day 2. In fact on Day 2 learners were also required to complete a fill-in-the-blanks task. Notwithstanding, young learners’ attitudes as to such tasks may also apply to the attitudes and subsequent performance of our 3rd-graders in our interpretation and production tasks. Shak and Gardner’s statistical analyses showed that learner enjoyment, ease, performance, and motivation on the GI task on Day 1 were higher in relation to the other 3 tasks. However, on Day 2 learner preferences on the GI measure decreased even though still higher or at the same level in relation to the other 3 tasks. More specifically, the mean performance value on Day 1 was 2.7 (std.: 0.5) and 2.5 (std.: 0.6) on Day 2. This drop may be attributable to the cloistered learning environment created by such and similar interpretation tasks since in such tasks the element of freedom is almost absent in the sense that there is usually only 1 correct answer and the learners are not free to respond without restraint. Such task design requirements may detrimentally affect learners’ enjoyment, ease,
performance, and motivation if utilized on more than 1 occasion, according to Shak and Gardner’s statistical analyses (p. 408).

The English tense system is considered by learners as one of the most difficult parts of English grammar to acquire (Scheffler, 2009, p. 9). It has been proposed that in cases of difficult rules explicit instruction is ineffective (DeKeyser, 2003, qtd. in Scheffler, 2009, p. 8). At the same time, Scheffler reported that respondents to his questionnaires regarded instruction on English tenses as very useful, in fact more useful than for any other area of English grammar (p. 9). Combining DeKeyser’s postulation and Scheffler’s data, leads to the assumption that implicit instruction might be the best way to teach the English past tense. In this case, both instructional purposes and learner requirements are met.

According to Ambridge, there are 3 major theoretical models that have attempted to account for the acquisition of the English past tense: (a) the single-route model, (b) the dual-route model, and (c) the multiple-rules model (Ambridge, 2010, p. 1497). Ambridge investigated the general underlying issue incorporated in such models, i.e. whether formal rules or analogies to stored exemplars govern language, and attempted to provide pertinent evidence. The author used young children as participants divided in 2 age groups: (a) 6- to 7-year-olds, and (b) 9- to 10-year-olds. These children were required to judge the acceptability of novel past tense forms, both regular and irregular, a research methodology often used with adult subjects. For Ambridge, such an age group, i.e. young children, constitutes a population sample most likely to procure the most valid confirmation or disconfirmation of the aforementioned theoretical models within the context of rule-based or analogically-based language acquisition. These children, who “are still learning the English past tense system” (p. 1498), can compliment the findings of similar studies on adults.
Zhuang posited that there are 2 elements that markedly differentiate languages: time and space (Zhuang, 2010, p. 1). If these 2 concepts are to shape teaching methodologies, linear approaches, i.e. movement from one grammatical rule to another in a step-by-step linear fashion, are not appropriate; rather, “cyclical interconnections” between rules should characterize the language learning process (p. 2). According to Zhuang, language learning requires abundant contextualized examples in sentences (p. 7). Zhuang, citing Little, continued that “for the implicit knowledge of grammatical rules, a developing mental lexicon is the necessity in the development” (Little, 1994, qtd. in Zhuang, 2010, p. 7).

Abrahamsson’s (2012) study belongs to a very different research paradigm and utilized a different experimental design; nonetheless, it concurs some findings and ensuing argumentation that relate to our study. First, Abrahamsson underscored the importance of focusing on learners’ errors (p. 197), an issue also addressed in our following discussion of Mourssi’s (2012) study. Second, the most relevant of Abrahamsson’s findings concerns the author’s tentative conclusion that young learners (early learners in Abrahamsson’s terminology) are capable of developing implicit learning capacities, while adults usually learn explicitly (p. 209 – 210). Such a postulation adds credence to our decision to expose our young learners to implicit teaching methodologies. Our decision is further substantiated taking into account that Abrahamsson was mainly concerned with whether early or adult (late) learners can ultimately achieve native-like proficiency. His summative results then further add credence to our research design, both as to the implicit instructional methodologies utilized and as to our choice of target age group.

A number of researchers have noted the problems EFL learners have with the acquisition of the English past tense and have conducted research to study this
phenomenon (Khatib & Alizadeh, 2012, p. 173). Khatib and Alizadeh pointed out that a very efficient way of investigating and teaching a tense such as the English past tense is through the use of picture-cued writing tasks; the combination of visual and verbal information to aid textual writing helps learners focus more on meaning rather than form (p. 180). Such a research design resembles closely our study where all exposure materials and testing tasks included both visual and verbal cues. It seems that such instructional tasks are an effective tool for helping learners overcome the problems they encounter with the English past tense.

One of the problems we faced in the main study concerned the classification and subsequent scoring of correct and incorrect responses. Mourssi’s (2012) analysis of the interlanguage (IL) stages Arab EFL learners go through towards mastery of the English simple past tense may offer valuable insights in this respect. Mourssi concluded that her data revealed the following interconnected developmental stages based on her categorization of Arab learners’ “non-target-like forms”:

1. base or present simple form
2. correct oral forms which were erroneous in written production
3. productive production of the –ed suffix
4. incorrect use of the auxiliary be with various verbal forms
5. erroneous use of number concord
6. incorrect blended forms
7. overgeneralization of various irregular past tense sub-rules
   (Adapted from Mourssi, 2012, p. 160 – 161)

The effect of memory on language acquisition is another field of study within which the English past tense has been examined (Lum & Kidd, 2012, p. 989). In this context, researchers have attempted to discover links between various proposed
memory systems, i.e. procedural and declarative memory, and the acquisition of the English past tense. Procedural memory is supposedly involved in the processing and acquisition of regular past tense forms, whereas declarative memory is believed to aid the processing and acquisition of irregular past tense forms (Ullman, 2001a, 2001c, 2004, qtd. in Lum & Kidd, 2012, p. 989). We are primarily concerned with the memory systems involved in the acquisition of the regular past tense, nonetheless, the effects of memory as to the acquisition of the irregular past tense might enable comparisons between regular and irregular past tense form acquisition.

Within this context, the relationships between memory systems and the past tense might shed light on our present findings. Our study included typically developing young children similar to the study of Lum and Kidd and contrary to the common research practice of using as target populations either adults with disorders of a neurodegenerative nature or children who have language disorders of a developmental nature (p. 991). Since Lum and Kidd included 5-year-old English L1 children and our study included 8-year-old English L2 children, a number of issues raised by Lum and Kidd may be applicable to our study as well. Kid and Lum emphasized the importance of targeting this age group since

“it seems that in typical development, 5-year-olds may be the oldest age group in which errors in regular past tense may appear but are not necessarily indicative of a language impairment or delay.” (Lum & Kidd, 2012, p. 992)

Exposure to the English past tense for the children in Kidd and Lum’s study was overwhelmingly richer than that of our young EFL learners. This means that, in our case, the errors as to regular past tense verbs may be an indication of linguistic impairment or delay.
The English past tense and, more specifically, the regular past tense has been reported as a possible indicator of language impairments, i.e. as a clinical marker differentiating typically developing (TD) young children from children with language impairment (LI) by Blom and Paradis (2013). After reviewing the relevant literature, Blom and Paradis suggested that “accuracy with tense has the potential to be a clinical marker in L2 English” (Blom & Paradis, 2013, p. 281). The 2 authors also discussed findings from numerous studies that “regular verbs differentiate between TD children and children with LI” (p. 282). Although we did not include irregular past tense forms in our study, Blom and Paradis’ discussion and findings as to the regular past tense might be applicable to our own research as well.

Such and similar arguments enhance the importance of our study, especially in respect with the errors our young learners produced. According to the 2 authors, irregular past tense forms exhibit a more straightforward acquisitional pattern dependent mostly on word frequency, whereas regular past tense forms are dependent on a host of other factors such as word frequency, as well as type frequency and variability. Even more so, regular past tense acquisition is dependent on the allomorph of the –ed suffix, i.e. /t/, /d/, and /id/. This type of allomorphic variation gives rise to various allomorphic schemata affecting regular past tense acquisition (p. 283 – 284).

The assessment measures utilized by Blom and Paradis were almost identical to those of our study. More specifically, to promote past tense production the 2 authors used sets of double pictures, i.e. in the first picture a child performed an activity and in the second picture the child had completed the activity. Although the past tense form elicitation technique differed between Blom and Paradis’ study and our study, there are profound similarities that validate cross-comparisons between the 2 studies. The fact that the participants in the 2 authors’ study were between 5 and 6
years old enables such comparisons and adds further credence to proposed or expected similarities between the 2 studies.

Despite the fact that we believe that the teaching methodologies in our studies were designed to foster implicit learning, it is necessary to fully account for whether this type of learning (and not explicit learning) was in fact promoted. Rogers, Révész, and Rebuschat (2016) conducted a study investigating explicit and implicit knowledge as to the acquisition of artificial Czech morphology. Without any direct links to our study, Rogers et al.’s discussion of incidental and implicit learning may serve our purposes. More specifically, the authors noted that

“when learning incidentally, participants may or may not become aware of the linguistic focus of the experiment. The key criterion when learning implicitly is that participants remain unaware of what is being learned throughout the experiment.” (Rogers et al., 2016, p. 782, italics in the original)

According to this account, incidental learning does not equate with absence of awareness; rather, absence of awareness is more correlated with implicit learning. In the absence of measures of awareness in our study, it difficult to postulate convincingly as to the exact nature of the type of learning fostered by our implicit methodologies. Just because our instructional treatments were labelled implicit is not a necessary and sufficient condition that our young learners learned implicitly, since no assessment of the actual type of learning fostered was made.

2.5.4. The acquisition of the English past tense: summary

Based on the aforementioned studies, a number of issues which should be addressed in our study were raised: (a) the age factor, (b) error patterns, (c) properties of base forms, (d) type and token frequency, (e) neighborhood friends and enemies, and (f) various theoretical models that have attempted to account for the acquisition of
the English regular and irregular past tense forms. Most studies cross-examined both regular and irregular past tense forms. To the contrary, our study focused only on regular past tense forms. The basic inference of these studies in relation to our study seems to be that the acquisition patterns in our study are more easily explained by attributes of the verbs / tokens of the verb form, rather than by our experimental conditions. For example, performance on a specific verb by our subjects might not be an attribute of our research design but might be based on attributes of these verbs per se, e.g. by their type and token frequency, or by morphophonological properties of base forms.

Conclusion

This chapter has attempted to describe the properties of the English past tense in relation to the linguistic background of our Greek L1 young learners. The major finding was that the acquisition of this tense by our young learners does not involve a straightforward application of the relevant Greek Aorist form, but, rather the acquisition of a new tense and aspectual system. On the other hand, the majority of studies reviewed suggest that a host of factors should be taken into account when explaining the results of our study, apart from correct enhanced performance. All this calls for a re-definition of our standard analyses based on performance patterns. In fact, it might be the case that erroneous performance patterns may provide more valuable insights than correct answers to our assessment measures.
CHAPTER 3:

TEACHING GRAMMAR IN THE L2

Introduction

This chapter outlines the theory and the research findings related with input enhancement (IE) and input flood (IF), input processing (IP) and processing instruction (PI), and structured input (SI). It includes 8 sections. In section 3.1. the theoretical foundations of input enhancement and input flood are presented. In section 3.2. the theoretical foundations of input processing, processing instruction, and structured input are outlined. In section 3.3 research studies within the input enhancement and input flood framework are reviewed. Section 3.4. includes a presentation of studies on input enhancement and input flood using as a target structure the English Past Simple Tense. In section 3.5. research studies within the processing instruction framework are reviewed. Section 3.6. includes a presentation of studies on processing instruction using as a target structure the English Past Simple Tense. In section 3.7. studies within the structured input framework are reviewed. Section 3.8. includes a presentation of studies on structured input using as a target structure the English Past Simple Tense. All the sections begin with a short introduction and conclude with a brief summary.
3.1. Input Enhancement & Input Flooding

3.1.1. Input enhancement & input flooding: theoretical background: introduction

This section presents the theoretical background of input enhancement. Input flooding is also included since textual and visual input enhancement techniques are very often examined together with or compared to input flooding. The presentation begins with a brief review of Sharwood Smith’s (1993) paper. Sharwood Smith was the first to introduce the term *input enhancement* providing both its definition as well as its theoretical foundations. It would not be an overstatement to note that almost all of the subsequent research and theorizing on input enhancement include references to Sharwood Smith’s (1993) paper. This section ends with Winke’s (2013) paper, tracing the path theorizing on input enhancement has followed in 30 years. Apart from the theoretical commentary of notable scholars and researchers, such as White (1998), Barcroft (2003), Leow, Egi, Nuevo, and Tsai (2003), Wong (2003), Rott (2007), Simard (2009), Hernández (2011), Park and Nassif (2013), and Winke (2013), this section also includes 2 meta-analyses, i.e. Lee and Huang (2008) and Han, Park, and Combs (2008). These meta-analyses are reviewed at length since they present the findings of many studies within the input enhancement framework and, on the basis of these findings, offer general theoretical explanations. The authors of these 2 meta-analyses offer comprehensive overviews of research on visual and textual input enhancement from a theoretical point of view. These overviews are an invaluable tool in understanding the studies on visual and textual input enhancement, and input flooding reviewed in subsequent section.
3.1.2. Input enhancement & input flooding: theoretical background

Sharwood Smith (1993) placed both the theory of SLA and its practical implications within the framework of (a) the selection of the input to be provided to the learner and (b) the appropriate timing to provide the pre-selected input (p. 165). Decisions concerning the selection of input and the appropriate timing for provision are of crucial importance since they entail theoretical, practical, as well as curricular constraints or choices.

According to Sharwood Smith, the focus on input presupposes 2 prerequisites: First, an *input processor* and, second, the *end product* (p. 167). This entails 4 distinct stages in learning, i.e. (1) the selection of the input, (2) the presentation – provision of this input, (3) input processing, and, finally, (4) the outcome of input processing. These processes are presented in Figure 1:

**Figure 1: Stages in learning based on input**

![Figure 1](image-url)

Selection → Presentation / Provision → Processing → Outcome of processing

Sharwood-Smith adopted the difference between comprehension and production, as well as the constructivist notion that knowledge (i.e. linguistic) needs to be compartmentalized into smaller units. For this to happen, input needs to be *enhanced* in order to become more salient and more easily learned. He employed the term *input enhancement* as a means of increasing *input saliency* (p. 176) and distinguished between *positive input enhancement*, i.e. making “more salient certain correct forms in the input” and *negative input enhancement*, i.e. flagging “given forms...
as incorrect, thus signalling to the learner that they have violated the target norms” (p. 177). The latter also involves corrective feedback (p. 177). Non-linguistic signals also operate as input enhancement, according to Sharwood-Smith.

According to White (1998), typographical enhancement is a more explicit input enhancement technique than input flooding, but less explicit than the provision of rules (p. 86). This is an important observation placing 3 input enhancement techniques on a continuum from less to more explicit. This continuum is presented in Figure 2:

**Figure 2: Continuum of IE techniques from more explicit to less explicit**

Provision of rules → Typographical enhancement → Input flooding

White’s continuum places input enhancement techniques within both a theoretical and a research framework. Theoretically, it was in line with Sharwood-Smith’s use of the term input enhancement (Sharwood-Smith, 1993, p. 176) and it placed input enhancement within the theoretical agenda on explicit and implicit learning / teaching. In actual research practice, it paved the way for the investigation and assessment of different input enhancement techniques.

Barcroft (2003) also adopted a broad view of input enhancement as including a number of techniques, such as typographical manipulations, input flooding, and explicit instruction. He distinguished between more “invasive” types of input enhancement and less “invasive” types of input enhancement (p. 49). According to Barcroft, the feature that distinguishes between the 2 is the extent of input alteration and input manipulation. The *more invasive* types of IE, such as explicit instruction,
modify the input to a great extent and involve more than mere comprehension of the input, whereas the less invasive types of IE, such as typographical manipulation, require only input processing on the part of the learners.

Barcroft made an important remark about the efficacy of various types of input enhancement techniques concerning the inter-relationship between the time allotted in input enhancement research designs and the learning outcomes procured. “In other words, when making pedagogical claims, the general principle that effectiveness = amount learned ÷ amount of time spent on learning should be considered” (Barcroft, 2003, p. 49, italics in the original).

This is a crucial observation that introduced the issue that input enhancement effects should not only be measured according to the learning outcomes, but, also, according to the time allotted to procure these learning outcomes. This is a crucial issue, especially in foreign language settings and sides with Williams and Evans’s (1998) postulation concerning “efficiency” (Williams & Evans, 1998, p. 155). Rate of acquisition is a constant requirement and concern any practicing teacher can testify to.

In their introduction on the theoretical framework behind input enhancement, Leow et al. (2003) pointed out that in most of the studies they had reviewed textually enhanced input had not procured the same learning outcomes as unenhanced input. They noted that textual input enhancement research studies had most often used a number of typographical enhancement techniques in order to increase the salience of targeted forms in the input. The underlying assumption was that this enhancement should focus learners’ attention to these forms in an implicit manner, thereby increasing noticing and subsequent processing in comparison to input that had not been enhanced (p. 2).
Wong (2003) examined the role of textual enhancement in SLA arguing that “(t)he goal of TE is to render more salient targeted forms of written input that are not perceptually salient so that learners will be more likely to pay attention to these elements” (p. 18, parenthesis added). Apart from the relevant literature review in textual enhancement L1 and textual enhancement L2 research, Wong discussed at length VanPatten’s Input Processing model of SLA and focused on the relationship between textual enhancement and attention to form. As to simplified input, Wong ascertained that it entails “using high frequency vocabulary … fewer pronouns and idioms … (and) … simple syntax, repetition and restatement of ideas” (p. 22, parenthesis added).

According to Rott (2007), input enhancement on vocabulary items as well as “intervention tasks” maximize vocabulary acquisition (p. 167). Further beneficial effects are procured if motivational and cognitive factors are taken into account. “These factors are subsumed in the Involvement Load Hypothesis, which predicts that the higher the involvement load, the higher the possibility of incidental vocabulary learning” (p. 167).

Lee and Huang (2008) carried out a meta-analytic review titled ‘Visual Input Enhancement and Grammar Learning’. As a general remark, Lee and Huang pointed out that treatment conditions including a variety of enhancement techniques exhibited minimal gains compared with treatment conditions including input flooding. Lee and Huang also proposed that the underlying principle of textual enhancement is to make target forms more perceptually salient for learners in order to facilitate subsequent processing based on Schmidt’s Noticing Hypothesis. They continued that typographical input enhancement long-term memory might profit by rendering forms in the input more perceptually salient.
In the majority of studies Lee and Huang reviewed visual input enhancement (VIE) was compared to input flooding, i.e. a more implicit condition than visual input enhancement, according to the authors. However, despite the contradictory overall findings, Lee and Huang cited Lee who had commented that learner comprehension “might be negatively affected by attention-drawing VIE techniques and (had) proposed that a balanced amount of intervention to teach L2 forms during reading would be needed” (Lee & Huang, 2008, p. 309, parenthesis added).

An important finding of Lee and Huang’s review concerned the proficiency levels of the subjects in the studies reviewed. More specifically, they noted the overall tendency of research within the visual input enhancement framework to include intermediate level learners. Lee and Huang attributed this tendency to the assumption that beginners might not possess the necessary reading skills to engage in visual input enhancement research, whereas advanced learners were probably already very competent as to the target structures investigated. This is assumedly a gap in the literature, since prior knowledge – irrespective of the means of assessment – is a confounding variable that might distort the results. In this respect, Lee and Huang noted that studies including learners with high levels of prior knowledge need to be interpreted with caution.

The most common visual input enhancement techniques Lee and Huang came across were boldfacing and underlining. Receptive knowledge measures and productive knowledge measures were the most common assessment measures of learning outcomes, despite the fact that Lee and Overstreet (1998) had previously “reported that VIE might have observable debilitating effects on the processing of meaning comprehension” (Lee & Overstreet, 1998, qtd. in Lee & Huang, 2008, p. 317).
Lee and Huang made a remark concerning the control groups, i.e. that in visual input enhancement research a “true control” group is usually a group “with reading conditions in which the form in focus occurred infrequently in intact texts” (p. 320). However, the issue of the exact amount of input required to implement such a control treatment was not specified by Lee and Huang. The issue is further complicated since, as Lee and Huang repeated, visual input enhancement treatments were most of the time compared to input flood treatments, instead of “no exposure conditions” (p. 321).

The whole issue of control conditions in SLA remains problematic to a certain extent. Many research designs, as Lee and Huang correctly pointed out, do not utilise control groups without any kind of exposure to target forms. As Lee and Huang implied, without however explicitly mentioning it, a true control group is one where participants receive absolutely no exposure to target forms. It seems justified to argue that a true control group is a group of learners who only participate in all the assessment measures employed, are absent during the treatment conditions, and, in the meantime, engaged with study-irrelevant tasks. In fact, many studies fail to report even the existence of a control condition, something which is necessary to make accurate comparisons. Not including a control condition is a limitation and any findings reported need to be interpreted with great caution. This, however, cannot justify the use of control groups that receive whatever type of input flooding, as well. In the best of cases, both situations are unsound methodologically.

Lee and Huang concluded by offering a number of recommendations. More specifically, they proposed that issues such as learner proficiency, prior knowledge, and developmental readiness need to be more accurately balanced. However, as previously mentioned, the first 2 issues are problematic issues, despite the fact that
Lee and Huang cited Lee who had proposed that visual input enhancement would expectedly present more beneficial outcomes for learners with some kind of prior exposure to target forms (p. 325). Finally, Lee and Huang called for studies that were to investigate the amount of visual input enhancement required to procure the desired outcomes.

Han, Park, and Combs (2008) carried out a meta-analytic review titled ‘Textual Enhancement of Input: Issues and Possibilities’. In contrast to Lee and Huang’s (2008) meta-analytic review on visual input enhancement, Han et al. did not statistically analyze the studies included in their meta-analysis, but, instead, chose to organize the finding of the studies reviewed in 7 distinct categories.

Han et al. placed the theoretical framework of textual enhancement (TE) on the work of Sharwood Smith (1991, 1993) who had introduced the term *input enhancement*, Schmidt’s (1990) *Noticing Hypothesis*, Long’s (1991) definition of *FonF*, Gass’s (1998) model of SLA, and VanPatten’s (1991) *Processing Instruction* model of SLA (Han et al., 2008, p. 597 – 599, references in the original). They reiterated that textual enhancement mainly aims at increasing the salience of the input in order to aid noticing and subsequent intake. They underscored Sharwood Smith’s distinction between “internally generated” input enhancement (i.e. generated by the learners themselves) and “externally generated” input enhancement (i.e. generated by the teachers or the researchers) (p. 598). Han et al. summarized the findings of their literature review as follows:

“(I)n some studies TE was highly effective in that it led to noticing as well as acquisition; in some it was moderately effective in that it led to noticing but not to acquisition; but in others, it did not appear to be effective.”

(Han et al., 2008, p. 600, parenthesis added)
In sum, Han et al. pointed out that the collective findings of the textual enhancement studies they had reviewed presented a rather “confusing picture” (p. 600).

Consequently, they chose to discuss the pertinent literature under 7 general umbrella topics:

1. The relationship between noticing and acquisition.
2. The effects of textual enhancement on comprehension.
3. The distinction between simultaneous and sequential processing.
4. The relationship between textual enhancement and enhanced targeted forms.
5. The relationship between textual enhancement and the prior knowledge of the learners.
6. The scope of textual enhancement and input flood.
7. The overuse of textual enhancement.

(Adapted from Han et al., 2008, p. 601 - 611)

For each of these topics that Han et al. elaborated on, one of their pertinent corresponding remarks will be alluded to and elaborated on:

1. “Input enhancement may . . . work in ways unforeseen by the enhancer”

(Sharwood-Smith, 1991, p. 131, qtd. in Han et al., 2008, p. 602, italics added).

This entails that there is no definitive way of assessing the exact degree of the effects of treatments like input enhancement or input flooding on individual learners.
2. It has not been decisively validated whether textual enhancement effects entail a trade-off with comprehension (p. 603). This entails a tentative detrimental effect of textual enhancement on the learners’ ability to comprehend the subject material, even though they may learn / acquire targeted forms.

3. This appears like a distinction between “incidental processing” and “intentional processing”. The former accords with Krashen’s Input Hypothesis, whereas the latter accords with Information Processing cognitive theory (p. 605).

4. The selection of targeted forms in applied SLA research has always been a debated issue. Possibly, there is no clear-cut answer, especially taking into account that textual enhancement research, despite prolific, has so far examined a small fraction of all the possible language forms available in a limited number of languages.

5. It has been suggested that textual enhancement is best-suited for learners who already posses some prior knowledge of targeted forms (p. 612). However, although this fact is highly appreciable in actual classroom practice, when conducting research it is highly questionable whether prior knowledge (irrespective of assessment measures) can be scientifically, as well as statistically, accounted for.

6. “Most of the studies in the TE literature exhibited varying degrees of conflation of TE and input flood, both being options of input enhancement” (p.
Textual enhancement increases salience perceptually (most often visually). Input flood increases salience through enhanced frequency of embedded target forms (p. 610). In this respect, Han et al. pointed out that

“problems may arise when input flood is used along with TE to control for the experimental text and at the same time to manipulate the control text—the so-called ‘unenhanced text’. For one thing, the results from the experimental group can be contaminated with ambiguity. For another, input flood may cancel out the effects of TE.”   (Han et al., 2008, p. 610)

Tentatively assuming that one of the utilities of a control group is to validate possible test-cumulative effects on acquisition, it seems advisable that in order to overcome the problems of the combined use of textual enhancement and input flood Han et al. referred to, pure control groups that do not receive any kind of exposure to target forms, i.e. continue with their regular research-irrelevant activities, are preferable.

7. This issue is further complicated since, as Han et al. pointed out, in the majority of studies that exhibited overuse of textual enhancement, input flood was also used. However, this suggested detrimental effect was tentative (p. 610 - 611). It may be more of a research-internal trait in the sense that, in the limited time available, researchers try to enhance target forms in whatever possible way is available, increasing frequency of target forms in treatment conditions being a common option.

Han et al. (2008) concluded by tentatively implying that within the framework examining explicit and implicit instruction there is a qualitative difference between
the 2, suggesting that more exposure time may be necessary for implicit instruction. They believed that this might produce even more beneficial learning outcomes than intensive and time-limited explicit instruction conditions.

Simard’s (2009) theoretical framework was based on the role and the various types of the implicit technique of textual enhancement in SLA. According to Simard, in textual enhancement the aim is to increase the perceptual salience of targeted forms with the underlying assumption that language acquisition is aided by an increase in focal attention. Also, Simard introduced in the literature of textual enhancement the term *intake* which had been previously used in the Input Processing and Processing Instruction research agenda.

Hernández (2011) discussed the role of explicit instruction (EI) and input flood (IF) in SLA. As to the role of explicit instruction, Hernández reported that, on the one hand, in a number of studies within SLA explicit instruction as a treatment condition has been identified as procuring significant outcomes. On the other hand, Hernández reported that a number of studies, especially within the input processing and processing instruction research paradigm, have borne out “that EI does not enhance acquisition when exposure to rich input is combined with meaningful, task-essential practice” (p. 162). As to the role of input flood in SLA, Hernández reported that the aims of input flood and processing instruction (an example of explicit instruction) coincide. More specifically, both attempt to increase the frequency and salience of target forms in the input. Input flood increases the frequency of target forms in order to increase noticing and acquisition. Hernández pointed out that input flood as a teaching technique is theoretically based on Schmidt’s *Noticing Hypothesis*, as well as on the *Frequency Hypothesis*. 
Park and Nassif (2013) after examining the research within the input enhancement framework, reported that “evidence has largely been mixed” (p. 1). Park and Nassif reviewed the literature concerning the interrelationship between input enhancement and target forms summarizing that, even though overall findings were not conclusive and more research was needed to be carried out, textual enhancement seemed to be more beneficial for forms of high communicative value, whereas low communicative value forms seemed less susceptible to textual enhancement treatment conditions.

Finally, after a thorough examination of the research on input enhancement, Winke (2013) listed a number of factors that have been proposed by other researchers as to the mixed finding of input enhancement on learning:

1. The input enhancement technique employed
2. The type of form used; its saliency and complexity
3. How relevant a given form is to the understanding of a text
4. Task instructions, also possibly affecting motivation
5. The interrelationship between text comprehensibility and the level of proficiency of the learners
6. Text length
7. Topic familiarity
8. Developmental readiness
9. Form frequency in the input
10. Amount of explicit information included with input enhancement.

(Adapted from Winke, 2013, p. 327)
3.1.3. Input enhancement & input flooding: theoretical background: summary

Despite the fact that more than 30 years have passed since Sharwood-Smith’s (1993) use of the term *input enhancement*, the basic principles put forth by Sharwood-Smith are still operant. The reason for this is that *input enhancement* is, in fact, a theoretical umbrella term that includes both explicit and implicit instruction techniques and methodologies ranging from input flooding to any type of instruction that manipulates the input offered to the learners. Therefore, Han et al.’s. (2008) aforementioned conclusion that the collective findings of the textual enhancement studies they had reviewed presented a rather “confusing picture” (p. 600, see also Park & Nassif, 2013, p.1) is not surprising taking into account the plethora of input enhancement techniques. However, within this broad framework research has mainly focused on visual and textual input enhancement and input flooding. In fact, visual and textual input enhancement techniques, and input flooding have often been confounded as research variables (Han et al., 2008, p. 610), despite the fact that they rest on different theoretical assumptions. As Han et al. correctly pointed out this conflation becomes even more problematic when input flooded texts are used in control conditions, which means that the participants in these control groups have been exposed to the target structure *contaminating* the results.

Theorising on input enhancement has also raised a number of important issues that apply to any teaching methodology. The first issue is *effectiveness* (Barcroft, 2003). In other words, the amount of time allotted for learning is an important parameter as to the assessment of the efficacy of any teaching technique. The second issue concerns the intricate inter-relationship between target forms and type of input enhancement techniques. In other words, the same teaching technique is
not suitable for all linguistic forms. This means that an important parameter of learning concerns the properties of the object of learning. Theorizing on input enhancement has been very sensitive as to this issue. The third issue concerns the proficiency levels of the participants. As aforementioned, Lee and Huang (2008) noted the overall tendency of research within the visual input enhancement framework to include intermediate level learners on the basis that low proficiency learners lack basic reading skills, whereas high proficiency learners possess high prior knowledge of target forms. The final issue concerns the age of the participants. More specifically, unlike research on PI which has mainly utilized young adult learners as participants, studies on input enhancement have often used younger school-age learners as participants. In other words, visual and textual input enhancement techniques and inputflooding have been more often adapted to different age groups.

3.2. Input Processing & Processing Instruction

3.2.1. Input processing & processing instruction: theoretical background: introduction

Both theory and research on input processing and processing instruction could be summarized through the work of VanPatten. VanPatten (1984) provided the theoretical foundations of input processing. VanPatten and Cadierno (1993) was the first study to investigate the effects of the practical implementation of input processing, i.e. processing instruction. Lee and VanPatten (1995) comprehensively organized in a coherent whole the theoretical framework of input processing and processing instruction. VanPatten and Oikkenon (1996) introduced structured input as a separate instructional tool. It would not be an overstatement to acknowledge that all
and any research and theorizing within the input processing branch of SLA is but a re-examination of VanPatten’s work from a different perspective. VanPatten has (a) provided a coherent theoretical framework for teaching / learning, (b) conducted extensive research, and (c) paved the way for avenues of further research.

### 3.2.2. Input processing & processing instruction: theoretical background

Although the role of grammar instruction in SLA has been variously questioned, research and pedagogy are testifying to the opposite. More recently, the role of input has been examined in various guises, e.g. input processing, input flood. All these approaches standardly assume that there is a role for input in SLA and that grammar instruction directed at input is beneficial. Such assumptions contrast with Krashen’s theorizing on input. He suggested that input which is comprehensible is on its own sufficient and there is no need for grammar instruction whatsoever (e.g. Krashen, 1985, p. 2).

IP is the first step of an extended input-based model of SLA (see VanPatten, 1993; VanPatten and Cadierno, 1993; Lee and VanPatten, 1995; VanPatten, 1996; 2000; 2002) that will be presented further along. In IP learners receive the linguistic data they are exposed to, i.e. input, while they are making form-meaning connections assigning syntactic categories to the lexis they understand. Input is converted to intake, which is the subpart of input that is in fact processed. Learners extract form from the input while their attention is on meaning and this usually takes place during actual communication. Grammar instruction is focused on the input, which is manipulated in appropriate ways to foster the desired results.
In order to comprehend IP, one must not neglect the distinction between *input* and *intake*. This is a crucial element of IP, which examines ways of converting input to intake. Input could be defined as the totality of language a learner is exposed to. Intake is a subpart of that language, the part that the learner pays attention to and assigns to working memory for further processing. In the words of Corder who originally employed the term, “input is what goes in, not what is available for going in, and we may suppose that it is the learner who controls this input, or more appropriately his intake” (Corder, 1981, p. 9, emphasis in the original). VanPatten adopted a slightly different approach to intake recognizing it as “the subset of filtered input that serves as the data for accommodation by the developing system. It is the input that has been processed in some way by the learner during the act of comprehension” (VanPatten, 1996, p. 10). By recognizing the necessity that input become intake, IP is in sharp contrast to laissez-faire anti-grammarian approaches, i.e. the *zero option*, which stress that input alone can suffice for acquisition processes. IP presupposes that learners are not capable on their own of assimilating the totality of input and, consequently, there are limitations to what the learner can attend to.

The concepts of *attention* and *noticing* are central in IP and contemporary SLA theorizing. Schmidt believed that instruction leads learners to notice preselected forms when they occur in communicative input (Schmidt, 1993, p. 26; 1994, p. 17). Paying attention to forms when these are heard or seen is considered a prerequisite to the acquisition of these forms according to a number of researchers. This is in line with VanPatten and Cadierno’s (1993) model of SLA (Lightbown, 1998, p. 184). However, there are limitations to what a learner’s mind can process. In other words, attention is capacity robbing and if the amount of stimuli operating exceeds available resources, then the cognitive system becomes overloaded and blocked, not allowing
the incorporation of further input. This applies especially to adult second language learners for whom the input may sometimes seem incomprehensible, barring them in this way from noticing it and, subsequently, acquiring the relevant forms. In sum, “(w)hen the task demands exceed the available resources, then both storage and computational functions are degraded” (Just and Carpenter, 1992, p. 4, parenthesis added).

Nevertheless, attention and noticing on their own are not necessary and sufficient conditions for input to become intake. In order to become intake a form needs, firstly, to be detected. This means that specific information must first be selected or engaged on (Tomlin and Villa, 1994, p. 192). Simply paying attention to various pieces of information is not enough for learning. What is necessary is detection of the various data in the input during IP. This will ensure that these specific bits of information will become intake (VanPatten, 1996, p. 16).

Another important factor that affects the acquisition of a form and has central status in IP is communicative value. The relative communicative value of a form can attract or disengage attention to it and can likewise aid or obscure the process of detection. The term “refers to the relative contribution a form makes to the referential meaning of an utterance” based on two features: (a) inherent semantic value and (b) redundancy (VanPatten, 1996, p. 24; 2000, p. 298). Forms which have both semantic value and are not redundant are high in communicative value and tend to be acquired easily, e.g. English -ing (present participle). This happens because they can be more easily noticed and detected, and thus become intake, something that affects their acquisition (VanPatten, 2002, p. 760).

The connection of form and meaning is one of the preferred outcomes of IP or, to be more specific, any contemporary type of grammar-oriented instruction. What
this means is that learners beginning from interpretation must first learn how to detect a form in the input and decode it assigning the appropriate meaning value to it. Once they have mastered this process, the next step is to produce the form in the appropriate circumstances, in order to satisfy their communicative needs. It is important to state that most often learners move from comprehension to production and not the other way round. This justifies the enhanced role of input in the process, but is certainly not intended as a negative evaluation of the role of output. VanPatten limits the creation of form-meaning connections to what takes place during interpretation. For this researcher “(f)orm-meaning connections refer to the mapping that the learner’s internal processors make during comprehension between referential real-world meaning and how that meaning is encoded linguistically” (VanPatten, 2000, p. 297).

In such a model, form-meaning connections affect both aspects of SLA, i.e. both input and output.

VanPatten proposed the formulation of a number of principles of IP:

“P1. Learners process input for meaning before they process it for form.

P1a. Learners process content words in the input before anything else.

P1b. Learners prefer processing lexical items to grammatical items (e.g., morphology) for the same semantic information.

P1c. Learners prefer processing “more meaningful” morphology before “less” or “nonmeaningful” morphology.

P2. For learners to process form that is not meaningful, they must be able to process informational or communicative content at no (or little) cost to attention.

P3. Learners posses a default strategy that assigns the role of agent (or subject) to the first noun (phrase) they encounter in a sentence/utterance. This is called the first-noun strategy.

P3a. The first-noun strategy may be overridden by lexical semantics and event probabilities.

P3b. Learners will adopt other processing strategies for grammatical role assignment only after their developing system has incorporated other cues (e.g. case marking, acoustic stress).
P4. Learners process elements in sentence/utterance initial position first.

P4a. Learners process elements in final position before elements in medial position.”

(VanPatten, 2002, p. 758)

Beginning from P1, it posits that the first thing learners do when they encounter input in the second language is try and decode the message. This tacitly assumes that form and meaning are two distinct features and, what is more, they are in an antagonistic relationship. Although the form / meaning distinction is not an easy issue to deal with in the sense that it appears they cannot be totally separated, it is obvious that what the majority of people do is communicate in order to primarily exchange messages, i.e. they are concerned with meaning. This applies to learners as well, at least those who are learning a language so as to communicate. Within this framework subprinciples P1a, P1b, and P1c make sense intuitively: if learners are concerned with meaning, then it is understandable that they will be attentive to content words that carry most of the propositional import in a sentence / utterance (P1a). In fact, they might be looking just for them. P1b seems substantiated by empirical evidence reporting that when learners are offered verbal inflections vs. adverbials of time they opt for the latter performing better when such adverbials are presented to them than when verbal inflections are in the input (Musumeci, 1989, qtd. in VanPatten, 1996, p. 23). Learners “focus, above all things, on words” (Williams, 2001, p. 338). This principle has been labelled the Lexical Preference Principle and is related to the main study in the present thesis which attempted, among other things to omit temporal expressions so as to aid learners to focus on the -ed morpheme in order to acquire the English Past Simple Tense. As far as P1c is concerned, it must be understood in the context of relative communicative value. The terms “more meaningful”, “less meaningful”, or “non-meaningful” are used to signify the
communicative value of a grammatical form. Forms of high communicative value are
the most obvious candidates for processing on the part of the learner.

P2 is an extension of P1. If attention is an effortful and limited capacity, and
accepting that learners will firstly focus on content words, then it is obvious that only
the remaining attentional capacity (if there is any left) will be oriented towards form.
This has direct pedagogical implications suggesting that meaning must be dealt with
first, i.e. become evident to the learner. Form, especially of the nonmeaningful type,
might have to be taught at a later stage. To the contrary, form that is meaningful is one
of the best candidates for instruction to begin with. Another conjecture of P2, together
with P1, is that since learners prefer processing lexis before grammar, maybe the
teaching of vocabulary should come before the teaching of grammar. Once learners
have acquired the necessary vocabulary, their attentional burden will be lightened
enabling the acquisition of meaningful first, then nonmeaningful grammar. Arguably,
P1 and P2 are inextricably entwined although it should be noted that VanPatten
reported no solid experimental evidence in support of P2 (VanPatten, 1996, p. 27).

A central tenet of IP is P3. VanPatten labelled this the First Noun Strategy. It
predicts that, for example, in causative sentences the learner will mistakenly assign to
the first noun phrase the role of actor. Let’s take the sentence ‘Eleni is having her car
repaired’ as an example. More often than not, English L2 learners interpret the
sentence as Eleni performing the action by herself. This means that they assign Eleni,
i.e. the first noun, the role of actor / agent. This is mistaken and IP is directed
specifically at altering learners’ reliance on this strategy. The same argumentation
applies to other structures as well, e.g. the passive. SLA research has been conducted
with the strategy in mind proving that it does exist (Ervin-Tripp, 1974; Binkowski,

Nevertheless, although it seems that P3 functions as a default setting, learners eventually develop the appropriate strategies. This is operationalized through P3a. In other words, despite the fact that learners begin with P3 on the way they grow accustomed to other ways of deriving meaning, more specifically, the semantics of lexis and event probabilities (VanPatten, 1996, p. 36), i.e. whether an activity can actually happen in real life. Considering the primacy of vocabulary already commented upon, this is not a surprising fact. English learners will start out with P3 based on the standard SVO English word order. This will account for the majority of communicative encounters they run into. At some point, they will come across nonstandard word orders, e.g. causative or passive structures, which will puzzle them at the beginning. They will firstly resort to the first noun strategy. Noticing the inadequacy of the strategy, they will subsequently rely on lexical meanings and real world probabilities. As their IL develops, they will eventually be able to process other low-communicative value cues such as case marking and acoustic stress (P3b) (for research that substantiates principles P3a and P3b, see VanPatten, 1996, p. 36 - 42).

Principle P4 does make sense intuitively but VanPatten reported only 2 studies related to it (Barcroft & VanPatten, 1997; Rosa & O’Neill, 1998, qtd. in VanPatten, 2000, p. 300). In all likelihood, it is again a matter of attentional allocation and memory. Elements in sentences / utterances are distinguished in 3 categories: initial position, medial position, final position. Probably P3 can be explained along the following lines: When we communicate, we are interested in the message of the interlocutor / writer. Waiting for his / her message, we seek for it right at the beginning. Silence is usually not a message (although it can be very convincing).
People are alert to the breaking of this silence and the incoming of messages. This means attention is focused on the breaking out of words from the speaker’s / writer’s mouth / pen. Once this occurs and after the initial propositional content is expressed our co-communicator will usually continue adding more information. This means that our initial attentional focus is (over)loaded with additional information and subsequently obscured. Elements in medial position are by definition of secondary priority since the initial ones will have consumed a part of limited attentional capacity.

This can explain why elements in final position are processed later than those of initial position but to explain why they precede those of medial position one needs to examine the variable of working memory. It is likely that the various pieces of information are stored in various parts of the brain. On the basis of attentional focus only, final elements would be the hardest to process. However, the impact on memory of these elements is unlike that of medial ones due to the fact that the former are followed by silence. This will distinguish them from the rest of the stream of information. Memory and attention will recognize this and assign them increased value.

PI is a type of instruction that aims at improving the quality of input that learners process so the amount of input that becomes intake is increased. To achieve this learners’ processing strategies are enforced or altered so that better form-meaning connections are accomplished (VanPatten, 1996, p. 55, 60). PI is implemented in the classroom in 3 consecutive steps: (1) explanation, (2) information about processing strategies, and (3) SI activities (p. 60).

The first step is the regular explanation offered to students when teaching form. It might involve rules on the formation of structures, an introduction to their
meaning, etc. This is nothing more than the presentation phase of the traditional PPP (Presentation Practice Production) model. In this respect, PI looks like a FonFS approach that preselects structures to be taught and makes use of existing practices.

VanPatten’s originality lies mainly in the second component of PI. Accordingly, learners are informed about processing strategies that might be incorrect. For example, during PI on the causative form learners are explicitly informed that for this particular structure the standard assumption that the first noun of the SVO word order is the actor / agent is not accurate. Instead, learners are notified that the structure presupposes that an actor other than the first noun performs the action.

All said, this information about processing strategies poses limits in scope. By emphasizing the fact that PI should alter existing processing strategies, PI is limited to structures that involve such incorrect strategies. The question arising is whether forms that do not involve erroneous processing strategies are equal candidates for PI (such forms are probably a majority). In fact, on these grounds, VanPatten himself has attacked other researchers (e.g. DeKeyser & Sokalksi, 1996; Salaberry, 1997) for not including attempts to alter learners’ processing strategies in their studies (VanPatten, 2002, p. 780 - 781).

Finally, PI attempts to practice the pre-selected forms by means of SI activities. The 2 essential characteristics of such activities are that (a) learners encounter forms while they are focused on meaning and that (b) at no point do they practice producing forms; they only process them (Lee and VanPatten, 1995, p. 102). Moreover, the input learners receive “has been manipulated in particular ways”, that is why it is called structured (VanPatten, 1996, p. 63). The main thing that theoretically distinguishes these activities from other similar ones is that learners do
not produce the structures at all; they are only called on to comprehend them. On the surface it looked like a rather strong claim but Lee and VanPatten added that learners should be led from input-oriented activities (i.e. comprehension - no production) to output-oriented activities (i.e. production) (Lee and VanPatten, 1995, p. 103). What this presupposes is that for these researchers the input and output phases are two distinct processes each in need of specific differentiated treatment.

Furthermore, VanPatten proposed a number of guidelines for the design of SI activities summarized as follows:

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.

(Adapted and summarized from VanPatten, 1993, p. 438 - 439)

The first impression that one confers upon closer inspection of the principles is that they do not constitute a novel pedagogical approach. To begin with, teaching one thing at a time is what the author of this paper has personally come across in the majority of textbooks. In addition, the fact that teachers should focus on meaning while practicing form, although not strictly adhered to by all theorists and materials designers, is an unquestionable tenet of most modern theories in SLA and their implementations. Furthermore, since learners are engaged in activities that practice
form they will do something with the input by definition. What is more, unless learners are studying for specific purposes (e.g. conversational or advanced literary English) there have yet to be seen approaches and coursebooks based on them that do not, for example, teach both speaking and writing. Nevertheless, in the majority of cases learners still receive input through the written mode, so VanPatten’s proposal seems well ascertained if one wants to maximize the effectiveness of instruction for acoustic, visual, or mixed types of learners. The fifth guideline does not constitute something new in pedagogy either. As a matter of fact, most activities in beginners’ coursebooks follow this rule and it should be noted that it is in contrast with purely communicative approaches which profess teaching discourse, i.e. communication, from the very beginning. Finally, keeping learners’ processing strategies in mind reminds one of the previous argumentations concerning the adaptability of PI to the teaching of other structures that do not involve erroneous processing strategies.

The same considerations apply to the division of activities and the types of activities offered. Lee and VanPatten divided SI activities in two major categories: (a) referentially oriented, which “use an immediate concrete reference to ascertain the truth-value of a sentence” (e.g. students choose one of two sentences that best matches a given picture) and (b) affectively oriented, “that ask for an opinion, require a personal response, or, in general, tap the student’s own world” (e.g. surveys) (Lee & VanPatten, 1995, p. 109). They also proposed that some examples of SI activities are: (a) supplying information, (b) binary options, (c) selecting alternatives, and (d) matching (p. 109).

Despite the fact that all this was not novel, it is apparent that VanPatten (1993) and Lee and VanPatten (1995) organized in a coherent whole an approach to grammar teaching that is exclusively input-based and merits attention. The activities were
provided within a framework that was soundly based on theory, although it could not claim novelty. The assumption of VanPatten and Sanz that SI can theoretically aid the internalization of grammar and at the same time takes learners’ psycholinguistic processes into account (VanPatten & Sanz, 1995, p. 171) is probably a valid one.

In theory, even though there are a number of objections, the overall framework does make sense as an input-based paradigm aiming at the conversion of input to intake. It is also psycholinguistic in nature in that it focuses on learners’ internal processing mechanisms and manipulates input accordingly. This is a type of instruction that can both aid in the acquisition of certain grammatical forms and explain the process of SLA.

3.2.3. Input processing & processing instruction: evolution

3.2.3.1. Input processing & processing instruction: evolution: introduction

The present section attempts a re-examination of the underpinnings of Input Processing (IP) and Processing Instruction (PI). Discussion deals with the background and the general theoretical framework within which IP is understood by the present researcher and uses as points of reference (a) the VanPatten (1984) theoretical paper which is the basis of the subsequent IP and PI literature, (b) the VanPatten and Cadierno (1993) study which opened the way to subsequent research on IP and PI, (c) Cadierno’s (1995) research on the Spanish Past Tense carried 20 years before the main study in the present thesis which uses as a target structure the English Past Simple Tense, (d) VanPatten and Oikkenon’s (1996) study, a turning point which marked a shift of focus from PI to structured input (SI), (e) Lee and Benati’s (2007)
studies incorporating input enhancement (IE) within the framework of SI, and (f) Marsden and Chen’s (2011) study which attempted to investigate whether referential or affective activities are the causative variable for the reported benefits of SI. All in all, this introduction also attempts a re-evaluation of IP and PI on the basis of a present-day analysis tracing the path IP and PI have followed since VanPatten’s (1984) paper more than 30 years ago.

3.2.3.2. Input processing & processing instruction: evolution: past & present

To begin with, contemporary teaching has been greatly influenced by computer information processing models. Information / input is loaded in the memory system of the computer, the necessary processing takes place, and the computer produces the required output. Although teaching and learning originate in antiquity, most present-day theorizing, research, and teaching / learning implementations view the learner as an input processor. Processing occurs in the human brain and the central nervous system, and the desired result / output emanates. It would not be an overstatement to acknowledge that the vast majority of the contemporary literature on learning is but an effort to describe these internal processes and present language learning in the format of logical, theoretical explanations.

Present-day research is incomplete without computer-aided statistical analyses. Teaching effects and learning outcomes need to be axiomatically measured accurately through assessment which is continually becoming more and more adaptable to computer-friendly environments. One needs to bear in mind that assessment transforms the teacher (or the computer) from the learner’s assistant, friend, and counsellor, to the latter’s judge. “The roles however of friend and judge
are, as we know incompatible” (Kapsalis, 1996, p. 554, personal translation from the Greek original).

An important theoretical postulation with noteworthy implications that summarizes the whole framework of IP as an instructional treatment is the conclusion that processing strategies (i.e. the First Noun Strategy) cannot be abandoned by learners after L1 acquisition, and, consequently, affect L2 (or L3, L4, etc.) learning (VanPatten, 1984, p. 66). The entire IP literature and PI as a teaching method is best understood under this prism. It sides on the Critical Period Hypothesis (CPH) (Lenneberg, 1967), assuming that L1 is in most cases acquired naturally early in childhood. It seems that PI is based on the assumption that L1 and L2 are not learned / acquired in the same way and that the L1 will always affect L2 learners’ L2 interlanguage systems. Such a state of affairs has profound implications in that PI, on the whole, does not conform to naturalistic learning methodologies, but it aims to improve L2 classroom or classroom-like environments. The natural context then of IP is cognitive-constructivist in the sense that language is not regarded in a holistic manner, but as consisting of internal rules (and subsequent learning strategies for learning these rules) that can be pinpointed and implemented in appropriate teaching formats.

In accordance with cognitive-constructivism, input needs to be structured in a way that takes into account that “learners need access to input that is communicatively or meaningfully oriented” (VanPatten, 1992a, p. 55) or, in different terminology, input that is “comprehensible and meaningful” (VanPatten, 1992b, p. 25). PI is solidly placed in the realm of instructed / classroom second / foreign language acquisition and considers non-classroom instruction as somehow deficient (p. 24). More to the point, PI is based on the theorem that there needs to be a materials designer / teacher /
educator to structure the input (cognitive scaffolding) considering all the while learner psycholinguistic readiness (p. 24). IP is not a laissez-faire approach. To the contrary, it aims at finding the optimal way to teach the carefully prepared input taking into account the learner. Furthermore, it assumes a teacher (or a computer) and a learner. The link between the 2 is the input. Input within the learners’ zone of proximal development (ZPD) (Vygotsky, 1962) and meticulously structured by the teacher.

In this respect, PI is nowhere near Scrivener’s “jungle path” lesson type which is “evolutionary” and might lead to “powerful personal insights”, ostentatiously learner-need centred (Scrivener, 1994, p. 34 - 37). Scrivener’s jungle path is indeed the apotheosis of self-/ other-discovery, and freedom. An overtly self-confident and experienced teacher is a necessary prerequisite. Jungle path is cognitive-constructivism’s bane if one considers that it is impossible for the teacher “to state the lesson’s objectives until after it has finished” (Scrivener, 1994, p. 35).

On the other hand, PI is teacher-centered as there is a clear theoretical (schematic) framework (e.g. VanPatten, 1984, p. 25 - 26), clearly-defined guidelines for teaching (e.g. VanPatten, 1993, p. 438 - 439; VanPatten, 2004, p. 14, 18), along with appropriately structured activities (e.g. Lee and VanPatten, 1995, p. 109), and an overt or covert explicit motive to learn. Nothing is left to chance; everything is pre-planned beforehand. Measurable objectives are meticulously evaluated through interpretation and production tasks which are assessment tests. The teacher is the implementator and the learner the processor. The objectives are clearly and easily assessed and the processing is linearly / sequentially arranged from left to right. VanPatten’s classical by now figures are enlightening:
Figure 3: Processes in SLA

1                           2                                                3
Input ➔➔➔➔ intake ➔➔➔➔ developing system ➔➔➔➔ output

Figure 4: Traditional grammar instruction

1                           2                                                3
Input ➔➔➔➔ intake ➔➔➔➔ developing system ➔➔➔➔ output
            ↓
            ↓
focused practice

Figure 5: Processing grammar instruction

1                           2                                                3
Input ➔➔➔➔ intake ➔➔➔➔ developing system ➔➔➔➔ output
            ↓
            ↓
heuristic devices
(e.g. interpretation strategies)
focused practice
3.2.3.3. Input processing & processing instruction: evolution: research

VanPatten and Cadierno (1993) is, chronologically, the first study in the IP and PI literature. VanPatten and Cadierno (1993) examined possible effects of PI and Traditional Instruction (TI) using as a target structure object pronouns in Spanish. The strategy the researchers sought to alter was the First Noun Strategy. Learning outcomes were assessed through interpretation, distractor, and production tests. The results from these tests showed that the processing instruction (PI) group performed significantly better than the traditional instruction (TI) group and the control (C) group as to the interpretation of the target structure. As to the production of the target structure, the differences between the PI group and the TI group were not statistically significant and both improved in relation to the C group.

Cadierno (1995) reported results similar to those of the original VanPatten and Cadierno (1993) study. She investigated Spanish Past Tense verb morphology, comparing 3 groups: PI, TI, and C. She examined the Lexical Strategy. Learning outcomes were assessed through interpretation and production tests. The PI group improved significantly in both tasks, whereas the TI group improved only in the production tasks. The PI group and the TI group did not generate significantly different results in the production task according to the statistical analyses reported. The C group showed no signs of improvement.

VanPatten and Oikkenon (1996) were the first to investigate structured input (SI) as the causative variable of the PI-centered research agenda. They investigated
whether the explanation phase or the SI activities were responsible for the results of previous studies. VanPatten and Oikkenon concluded that SI was the causative variable responsible for the learning outcomes in that and previous similar studies.

Lee and Benati (2007) after conducting a series of studies including a variety of target forms concluded that their results constituted evidence that it does not matter whether SI activities are presented using input enhancement formats or not, but that the SI activities were the causative variable of the findings in their numerous studies (Lee and Benati, 2007, p. 125).

Marsden and Chen (2011) examined the relative effects of referential and affective activities included in PI materials. The target structure they examined was the Past Simple Tense -ed verb inflection. Four groups were formed: (a) Referential + Affective, (b) Referential only, (c) Affective only, and (d) C. The results of the statistical analyses conducted on the various tests showed that even though learning was enhanced by the referential activities, participants did not show improvement after exposure to the affective activities (p. 1079). Furthermore, their analyses implied that referential activities might exhibit traits of explicit knowledge learning (p. 1084).

3.2.3.4. Input processing & processing instruction: evolution: final remarks

In practical terms, within the IP research paradigm, the teacher, gradually, has become a facilitator limited to providing the pre-provided and pre-prepared structured input practice activities. Comprehension and production have become assessment tools. Both practice and production can be readily delivered through computers further limiting the future role of the traditional teacher who is now required to become an educator / researcher / materials designer / implementator. In fact, by
making SI referential and affective activities the causative variable of IP and PI, PI as a teaching tool has been decontextualized and IP has become something like a psychological construct which the individual learner is called to decipher and will be, subsequently, assessed on in front of a computer screen. The gradual progression and deconstruction of the original PI implementation to its various elements is a natural evolution of the cognitive-constructivist motivation to enhance learning outcomes. This, of course, presupposes a change in perspective. It calls for a redefinition of the traditional roles and notions of teacher and learner. Despite superficially learner-centered, all this is a shift of emphasis from the learner to the outcomes of learning.

The shift from output to input signals a replacement of the learner’s behaviour as the focus of attention with input in the form of focused practice becoming the focal point. Traditional associationist approaches have always considered observable (linguistic) behaviour, i.e. output, as the crucial element, according to the implications of the previously presented Figure 4. However, apart from Audiolingual behaviourism, highly successful in World War II training U.S.A. military personnel (Richards & Rodgers, 2001, p. 50), there is no reason to believe that even negatively criticized methods such as Grammar Translation, or Socrates’ dialectic (wrongly accused as logocopic) have ever ignored the necessity of appropriately structuring input. IP focuses on input with the belief that traditional output-based practice, i.e. traditional instruction (TI), is like “putting the cart before the horse; the learner is asked to produce when the developing system has not yet had the relevant intake data” (VanPatten, 1993, p. 436). Nevertheless, it is learned output, be it in the form of interpretation as to circling picture A or B, or in the form of oral or written production, that has always been and still is the object of assessment.
The underlying assumptions of IP and PI are far from Long’s definition of “focus on form” (Long, 1991, p. 45 - 46). In this respect, PI is Herbartian-oriented, a clear example of “focus on forms” in Long’s (1991) terminology. Only Scrivener’s “jungle path” (Scrivener, 1996, p. 34 - 37) is an unambiguous paradigm of “focus on form”, outspokenly “incidentally”-driven, once again in Long’s (1991, emphasis in the original) terminology. Similar considerations apply to all other forms of input enhancement, including input flooding and structured input. In fact, it could be argued that the use of the term implicit to describe such methods and techniques is, to a certain extent, misleading. This is so since the learners are indeed required to learn implicitly, but the input has been pre-selected and explicitly designed and manipulated by the materials designer. In other words, input enhancement, input flooding, and structured input are attempts to force learners learn in an implicit way on the basis of input which has been designed with an explicit aim / motive on the part of the materials designer. Learners are most of the times unaware of these aims and motives. Input enhancement, input flooding, and structured input are nothing more than tools that aid learners discover on their own the rules in the input, rules that have been pre-selected, meticulously prepared, and covertly presented by the materials designer. At least, Scrivener’s jungle path left the field open for the teacher to make his / her own choices.

In other words, although input enhancement, input flooding, and structured input seem not to incorporate the presentation phase of the classical Herbartian PPP model, this phase has been subsumed by the meticulous preparation of the materials to be taught. It seems that whereas in the teaching methodologies of the past the element of freedom was missing on the part of the learners, modern methodologies have
eliminated the element of freedom from the teacher, too. Consequently, freedom of choice has moved from the teacher to the materials designers.

3.2.4. Input processing & processing instruction: theoretical background: summary

VanPatten and Cadierno’s (1993), VanPatten and Oikkenon’s (1996), Lee and Benati’s (2007), and Marsden and Chen’s (2011) theorizing and research findings signal a progressive shift of emphasis from PI to SI to SI & IE and, finally, to the decomposition of the components of SI. The most noteworthy contribution is to be located in the VanPatten and Oikkenon (1996) study. By excluding the explanation phase of PI, the Presentation phase of the (Herbartian) PPP model was no longer included making practice the causative variable of instruction and production the assessable outcome.

3.3. Input Enhancement & Input Flooding: Research

3.3.1. Input enhancement & input flooding: research: introduction

The following section includes research carried out within the theoretical frameworks of visual and textual input enhancement, and input flooding. The decision to review studies that have incorporated input enhancement and input flooding in their research designs in the same section is based on the fact that in most cases the 2 teaching methodologies have been investigated together either as separate variables or combined. More specifically, in most studies visual and textual input enhancement techniques are either embedded in input flooded texts or cross-examined with texts
that include input flooding. Furthermore, in a number of cases textually enhanced conditions are compared with baseline control conditions that include the same texts with an input flood on the target structures.

In fact, although in theory the 2 methodologies are based on different theoretical foundations, in actual research practice it is very difficult to draw a dividing line between the two. As aforementioned, visual and textual input enhancement techniques are considered more explicit and obtrusive than the more implicit methodology of input flooding (see e.g. White, 1998, for a continuum between the two). A notable similarity between the 2 teaching methodologies is that target forms in the input are pre-selected and the texts are manipulated to enhance noticing and subsequent acquisition. However, the main difference between the 2 lies in the fact that whereas in textual and visual input enhancement noticing is explicitly fostered and the emphasis is on designing the most appropriate technique to promote acquisition, in the case of input flooding there is a shift of emphasis to the learners themselves who are expected to discover the underlying L2 patterns on their own. In other words, visual and textual input enhancement borders with explicit instruction, whereas input flooding is based on premises like naturalistic learning conditions and Krashen’s (1985) input hypothesis according to which simple abundance of input is a basic prerequisite of acquisition. All said, the issue of the amount of input required to foster acquisition remains unresolved even though in actual research practice researchers are required to accurately balance the amount of input offered in each case.

3.3.2. Input enhancement & input flooding: research: studies
Trahey and White (1993) investigated the effects of input flooding (IF). The target structure selected was adverb placement in L2 English. The theoretical framework of Trahey and White was based on the UG principles and parameters framework. Trahey and White’s main objective was to answer the question whether the UG principles and parameters framework assuming positive evidence as the causative variable for L1 acquisition could be applied to L2 acquisition, “or does negative input also play a role (in L2 acquisition)?” (Trahey and White, 1993, p. 182, parenthesis added).

The final participant pool consisted of “an additional” (p. 187) 54 5th-grade French L1 children learning English in Quebec, Canada. The participants formed 1 group: input flood. It needs to be noted that Trahey and White’s study was carried out as a follow up of White’s studies (White, 1991a; 1991b, qtd. in Trahey & White, 1993, p. 185 - 187), hence the use of the word additional to describe the subjects in their study. Learning outcomes were assessed through a grammaticality judgement / correction task, a preference task, a sentence manipulation task, and an oral production task. The combined results (i.e. Trahey & White, 1993; White, 1991a; White, 1991b) lead Trahey and White (1993) to conclude that input flooding, i.e. “positive L2 input”, had a beneficial effect as to L2 learners’ interlanguage systems concerning adverb placement. Nonetheless, this input flooding was not adequate to eliminate “forms that are permitted in the L1 but not the L2” (Trahey & White, 1993, p. 201).

Alanen (1995) investigated the effects of input enhancement in relation to the presentation of rules. The target structures selected were linguistic structures in semi-artificial Finnish. The theoretical framework of Alanen was input enhancement and explicit information. Alanen focused on a number of issues, such as the distinction
between *input* and *intake*, the role of noticing in language learning, the role of perceptual salience, and the contradictory learning outcomes of implicit and explicit learning conditions (p. 262).

The final participant pool consisted of 36 English L1 students in a university in the USA. The participants formed 4 groups: Enhance group (targets of learning in italics), Rule group (explicit rule information), Rule and Enhance group (explicit rule information plus targets of learning in italics), and Control group (simple provision of text). All the participants received the same 2 texts as reading material. Learning outcomes were assessed through a number of tests: sentence completion test, comprehension test, word translation test, grammaticality judgement test, rule statements, and think-aloud protocols. The results of the detailed statistical analyses procured a number of findings. Overall, the 2 groups that received explicit information, i.e. the Rule group and the Rule + Enhancement group, produced enhanced learning outcomes compared to the other 2 groups that did not receive explicit information, i.e. the Input Enhancement only group and the Control group. In fact, there were no “clear-cut differences” in-between the 2 sets of groups (p. 288).

Alanen attributed the non-effects of input enhancement to a number of factors the most pertinent of which were (a) that “the method of highlighting (i.e. italics) may not have been perceptually salient enough”, and (b) “insufficient attention or too shallow processing” (p. 288 parenthesis added). The author also reported a number of other findings, as well. She reported that explicit information (EI) led learners to make specific errors and might have made learners *over-rely* “on L1 routines” (p. 289). Also, the RE (rule explanation + explicit information) group performance on the endings in the grammaticality judgement test was extremely meagre. Alanen attributed this to differences between the L1 and the target language, as well as to
individual learners’ IL systems. Finally, the performance of the Control group in Alanen’s study exhibited a number of signs pointing tentatively to enhanced learning outcomes. Alanen attributed such signs to the role of prior knowledge and expectancies of learners, as well as individual differences (p. 292).

Alanen’s (1995, p. 292 - 293) remarks concerned both the Control group and the role of input enhancement. More specifically, the author pointed out that the partial acquisition gains of the Control group could have been attributable to the role of individual differences in SLA. In fact, Alanen’s Control group of learners were exposed to the 2 texts the other 3 treatment groups were exposed to. According to Alanen, “even a small amount of help in focusing the learners’ attention on learning targets may help to level some of the differences among individual language learners” (p. 292 - 293). However, this means that Alanen’s Control group was not a true control group in the sense that it did receive exposure to target forms. Alanen’s subsequent claim that input enhancement might have “influenced learners’ approach to the task without them being aware of it” (p. 293), referring to the gains of the Control group, is probably in the right direction, but also further highlights the fact that the Control group were in fact exposed to the target forms.

Lorch et al. (1995) investigated the effects of typographical cues on text memory. In 2 experiments they tried to examine how reading passages is affected by various types of signalling, the term they use to refer to enhancement. Lorch et al. wanted to determine (a) whether typographical cues affect the recall of cued material, (b) whether typographical cues affect the recall of non-cued material, and (c) the way typographical cues affect readers’ on-line attention (p. 51).

Lorch et al. pointed out the fact that over-use of input enhancement techniques by an author may have a detrimental effect on learning since, either the author has
failed to emphasize the material focal attention is required for, or the readers are overwhelmed by over- emphasising (p. 52). After a review of the relevant literature, Lorch et al. concluded that “a frequent characteristic of experiments failing to demonstrate typographical cuing effects is that a large proportion of the text content is signalled” (p. 52).

Lorch et al.’s observation is probably in the right direction since the purpose of input enhancement is to render target forms more salient. Over-enhancing texts in this sense may distract learners’ attention and procure detrimental effects on learning. The same may be true of the simultaneous use of different enhancement techniques.

In Experiment 1, the final participant pool consisted of 124 university students from “introductory psychology courses and a sophomore-level course in Developmental Psychology” (p. 54). The participants formed 3 groups: No Underlining (control) group, Light Underlining group (5% of words underlined), and Heavy Underlining group (50% of words underlined). Moreover, “in the two conditions containing typographical cues, half of the target statements were underlined and half were not” (p. 52). Lorch et al. summarized the results of Experiment 1 by pointing out that when enhancement techniques are overused, learners simply ignore them, whereas judicious use of such techniques aid the learners’ task. Lorch et al. also pointed out that whereas learners’ memory operations for enhanced material are assisted, their memory operations for unenhanced material are not affected (p. 57).

In Experiment 2 the final participant pool consisted of 80 students “from introductory psychology courses” (p. 58). The participants formed 2 groups: No Capitalization (control) group and Light Capitalization group. There were an equal number of subjects in each group. Lorch et al. summarized the results of Experiment 2
as indicating that, once again, even though enhanced material produces prolonged attentional focus for this enhanced material, this does not affect attentional focus on unenhanced material (p. 61).

Combining the findings of Experiment 1 and Experiment 2 of their study, Lorch et al. assumed a number of hypotheses concerning the relationship between typographical enhancement and text memory. More specifically, they assumed that typographical “cues do not divert a reader’s attention from unsignaled content to signaled content, nor do they produce a general increase in attention to signaled and unsignaled content. Rather, typographical cues appear to affect processing only of the signaled content” (p. 63). Moreover, judicious use of cues by the author coupled with cues that aid readers achieve reading purposes, assist reader attention to the cues during the process of reading (p. 63). Finally, while enhanced information enables the reader to pay more attention to the material by reducing the speed of reading, the “processing of unsignaled content is unaffected (with the possible exception of a short-lived “spillover effect”)” (p. 63, parenthesis and quotation marks in the original).

White (1998) investigated why learners are sometimes unable to notice some target forms that are present but nonsalient in the input. The target structure selected was third person singular possessive determiners (PDs). White’s final participant pool consisted of 86 elementary school French L1 learners in Canada (exact grade or age not reported). They formed 3 groups: Group E + (Textual Enhancement + Input Flooding + extensive reading and listening), Group E (Textual Enhancement + Input Flooding), and Group U (Input Unenhanced + Input Flooding). All the texts used by White were modified stories, fables and poems for children for whom English was an L1.
White used different types of textual enhancement in the various activities utilized in order to augment the originality of textual enhancement techniques and ensure that learner attention was further focused on target forms, while, at the same time, she ensured that the enhancement was not rendering targeted forms extremely salient so as not to distract or irritate learners (p. 90). This final point echoes similar aforementioned arguments proposed by Lorch et al. who referred to possible detrimental effects of over-emphasizing target forms on learning (Lorch et al., 1995, p. 52).

White pointed out that sometimes comparison groups exhibit differential learning outcome in follow-up post-tests than in immediate post-tests. The causative variable for this may be that comparison groups “catch up” with experimental groups, or may be due to learners’ forgetfulness (p. 92).

Learning outcomes were assessed through a passage correction task, a multiple choice test, and an oral picture description task. White noted that “(o)nly the PD data collected during the picture description task will be discussed in this chapter” (p. 93, parenthesis added). The results of the quantitative statistical analyses focused on 3 domains: (a) frequency of use, (b) intralearner variability, and (c) accuracy (p. 94 - 97). Qualitative statistics were also included (p. 97 - 101).

White summarised her findings in the following way:

“Although accuracy ratios overall followed the predicted order, that is Group E+>> Group E>> Group U, the within-group variance cancelled out most of the predicted between-group effects at the two posttests.”

(White, 1998, p. 101)

Earlier on, based on the qualitative analyses, White had pointed out that the developmental paths followed by the learners in her study were not the same. A number of learners exhibited gradual progression, whereas some others progressed
fast. The latter ones either moved to developmentally earlier stages or remained unaffected at the delayed post-test. Finally, a small number of learners did not exhibit any gains whatsoever (p. 99).

Since the common element in all 3 instructional treatment groups was input flooding, it could be tentatively assumed, that the quantitative analyses and the qualitative analyses combined may lead to the conclusion that input flooding affects individual learners in a varied, differentiated manner. This is to be expected since this technique is highly implicit and unobtrusive; consequently, in the absence of any other type of enhancement, learners are left on their own to make sense of the input provided. It is likely then that, in the absence of explicit guidance, learners should respond to the exposure of input in idiosyncratic, non-uniform ways. On this basis, it is justified to assume that input flooding, as an instructional treatment technique includes an element of freedom in the sense that learners are free to respond in any possible way without being exposed to explicit teaching techniques.

In order to explain the results, White proposed that 1 factor may have been that the written tests (results of which were not reported in White, 1998) may have increased the salience of the target forms, i.e. that these “tests may have enhanced the target forms similarly for learners in all three of the treatment groups, thereby reducing the differences between them” (p. 102).

This is a problem in the SLA literature. It is an indisputable fact that the inclusion of assessment measures in a research design may, on its own, operate as a type of input flooding that may distort the results. In other words, the tests themselves may enhance target forms. There is no easy solution to this problem. The most obvious solution is to carry out multiple identical research designs on a variety of target forms in order to accurately assess the effects of an instructional treatment
technique. However, even if this were the case, so-called test-taking effects cannot be excluded. Consequently, as White correctly pointed out similar comments should apply to the majority of quantitative research carried out. However, this drawback can be eliminated to a certain extent by carefully designed qualitative free-response tasks including task-essentialness, i.e. the necessity of using target forms in order to carry out tasks (see e.g. Loschky & Bley-Vroman, 1993, p. 132).

White (1998) also added that the fact that learners had already come across the target form during previous instruction may have distracted learners’ attention to the extent required (p. 102). This is another problem in the SLA literature. The solution usually utilized is cut-off scores to ensure minimal prior knowledge of target forms. However, this is a quite problematic issue as discussed further along. The only real solution to the problem is to select young beginner learners with minimal exposure to the target language and, at the same time, select target forms that are far beyond these learners’ proficiency level. If, however, as in White’s (1998) case, learners have already been previously exposed to target forms in one way or another, this raises serious reservations about the validity of the findings, since there is no valid way of accurately assessing the real effects of an instructional treatment in the presence of prior knowledge.

According to White, another causative variable for the findings may have been the absence of any type of explicit input enhancement. She noted “that many of the learners in this study might have benefited from a more explicit type of enhancement” (p. 92). She added that whereas this explicitness may have benefited some learners, others “may have been more comfortable with the inductive approach used in this study and more able than other individuals to figure out the patterns in the input on their own” (p. 103).
White’s commentary on *explicitness* might also lead to the aforementioned assumption that implicit input flooding has differentiated, varied effects on individual learners. This variance is, to a certain extent, *undesirable* in statistics as high values in the measure of variance imply that a variable does not produce robust effects. However, it is in accordance with learner-focused pedagogical approaches which believe that interlearner and intralearner variability is not only desirable, but also *natural*. The fact that individual differences between learners might be interpreted as implying that some learners may benefit from more explicit treatment techniques while for other learners more implicit treatment techniques might prove more beneficial cannot be easily verified. Large-scale longitudinal cross-linguistic studies are needed to validly decide on the issue, especially in the case of implicit treatments.

White also noted that it is hard to decide whether input flooding, textual enhancement, or multiple test administrations were the causative variable of the enhanced performance, even though she admitted that input flooding was the common element of all 3 treatment groups (p. 103 - 104).

White’s final explanation was within a development framework including 8 stages in the acquisition of English PD rules. Accordingly, ‘U-shaped development’, or restructuring may have been the reason behind “the shift from correct use of a target feature to a developmentally earlier L2 feature and eventually back again to target like use” (p. 105).

This explanation assumes that learning / acquisition is not a straight right to left linear progression but includes also left to right *movements* as learners constantly restructure their interlanguages. The term *movements* is used deliberately at present instead of *regression* as it is viewed as a sign of progress in the sense that although it might be statistically assessed as a *detrimental effect*, it is, in fact, a sign of *trial and
error learning (see e.g. Thorndike, qtd. in Kapsalis, 1996, p. 238), or gestalt learning (see e.g. Wertheimer, 1959), or discovery learning - spiral progression (see e.g. Brunner, 1966). In order to validly measure such movements and learning in this sense, once again, large-scale longitudinal cross-linguistic studies are needed since very limited in time-span studies may lead to erroneous assumptions, i.e. a regression to a previous developmental stage may be misinterpreted as a detrimental effect on learning.

However, White (1998) attempted to explain restructuring as an effect of the memorization of chunks in the input flooded passages. This might have benefited learners “more memory-oriented than others” (p. 106). A second attempt to explain restructuring was also provided, namely that it was the outcome of the coding procedures utilized (p. 106). The former explanation is based on individual learner differences, whereas the latter explanation is based on characteristics of the assessment measures.

White concluded by arguing that her study lead to the assumption that “implicit F on F instruction may not be adequate in cases involving L1-L2 contrasts” (p. 106). According to White, positive evidence may not suffice and explicit treatments may prove more beneficial.

Williams and Evans (1998) compared the relative effects of (a) input flood + FonF instruction and (b) explicit instruction and feedback. The target structures selected were 2: participial adjectives and passives in English. The authors’ main aim was “to answer some preliminary questions about two related aspects of FonF research and instruction: the choice of forms for FonF instruction and the interaction between choice of forms and instructional treatment” (p. 139).
The participants were university students attending 3 intermediate English L2 writing classes at a university in the USA. Eleven learners in each class participated in the pre-test and post-test assessment measures and in the 2 treatment conditions. The participants formed 3 groups: Group I (input flood + explicit instruction and feedback), Group F (input flood + “a somewhat greater focus on form than a flood alone”) (p. 142) and Group C (control). “Participial adjectives were already used, though often inaccurately by the students” (p. 143). Learning outcomes for the participial adjectives were assessed through a grammaticality judgement task and a sentence-completion task. Learning outcomes for the passives were assessed through a more open-ended sentence completion task and a written cue-based production task. Furthermore, instructional treatment written assignments were collected as well as information from a dictogloss task.

As for the participial adjectives, there was “a significant difference between Group I and Group C, as well as between Group I and Group F, but not between Group F and Group C” (p. 148). As for the passives, in the narrative task “only the instructed and control groups’ posttest scores showed a small significant difference” (p. 148), whereas in the sentence completion task “both Group I and Group F showed significantly greater increases than Group C but were not significantly different from each other” (p. 148, 150). The qualitative data gathered from the dictogloss “were in general consistent with those from the more structured posttests” (p. 150).

Williams and Evans provided the following explanations for their findings (p. 151 - 154). First, individual readiness is a key factor, meaning that learners who already have some kind of learning on the target forms are most likely to improve irrespective of experimental condition. Second, there exists a differential effect of instructional types on different target forms. According to Williams and Evans, the
forms that are most susceptible to explicit instruction are the ones that learners misinterpret or have difficulty in analyzing. For such forms negative evidence seems to prove beneficial. Also, the authors assumed that for the passives a more explicit instructional treatment such as Processing Instruction might have been optimal, whereas for the participial adjectives, the input flooding was effective but not as effective as the explicit instructional treatment of Group I (p. 154).

Williams and Evans concluded that the crucial issue is “efficiency”. Even though learners who are ready may improve their learning skills irrespective of instructional treatment, instruction that provides a focus on target forms does increase speed of acquisition (p. 155). This final remark is of great importance, especially in foreign language learning settings, since, due to the limited time and input available, speeding up learning is highly desirable. Learners in such settings usually attend classes which operate as the sole providers of input. In the absence of external extra-curricular input, as in e.g. immersion settings, the prospects of efficiency and increase in the rate of acquisition are of great importance, especially since un-focused instruction is usually better in the long-run, but, at the same time, less speedy.

Leow (2001) investigated the effects of written enhanced input in SLA. The motivation of Leow was to assess more accurately written enhanced input learning outcomes based on online assessment tools. The target structure selected was the Spanish formal imperatives. The author’s theoretical framework was based on the roles of attention and noticing in SLA. After a review of the pertinent literature, Leow concluded that “failure to operationalize and measure the process of attention in second language acquisition studies” as possibly the most significant drawback in the assessment of learning outcomes in SLA (p. 497).
The final participant pool consisted of 21 college students attending a 1st-year Spanish program. Due to the fact that these participants had already been exposed to the target form, Leow used a series of elimination techniques to ensure that “only participants who had minimal knowledge or ability to recognize and produce in writing” the target form were included (p. 499). The participants formed 2 groups: Enhanced group and Unenhanced group. The enhancement techniques used were the following: omission of secondary information, underlining all verbs, and bolding the suffixes of targeted verb forms. Learning outcomes were assessed through a multiple choice recognition task (measuring intake), a fill-in-the-blank task (measuring written production), and a comprehension task (measuring comprehension), as well as a think-aloud protocol.

The statistical analyses conducted procured the following results. First, tentatively, written enhanced input promoted more comments on target forms. Second, no statistically significant differences as to reported noticing and consequent intake between the 2 groups were reported. Third, “highly abnormal” data distributions rendered impossible the assessment of delayed effects as to reported noticing and consequent intake between the 2 groups. Fourth, “highly abnormal” data distributions rendered impossible the assessment of the relationship between reported noticing of targeted forms and written production of targeted forms (p. 502). Finally, no statistically significant difference as to the comprehension of textual material between the 2 groups was reported.

Based on these findings, Leow argued that in his study noticing was unaffected by written textual enhancement. Moreover, he argued that mode of exposure may affect noticing in different ways. Also, he postulated that written enhanced input does not aid learners in processing L2 forms any more than
unenhanced materials do. Leow also scrutinized the think-aloud protocols of high outliers stating that they exhibited higher levels of awareness, implying that individual differences pay an important role in learning.

Leow concluded (a) by emphasizing the crucial role of the amount of exposure to target forms required to procure desired learning outcomes and (b) by arguing in favour of applying both quantitative and qualitative analyses in order to examine noticing in SLA (p. 506 - 507). As to the former conclusion, i.e. concerning the amount of input required, this is a highly debated issue without an easy explanation since it is influenced by individual differences, different learning styles, and the available time of exposure to the target language. As to the latter conclusion, i.e. the incorporation of both quantitative and qualitative analyses, even though qualitative analyses are unquestionably justified from a research perspective adding validity to any type of research agenda, they also open the issue of the amount of data that can be safely collected and compared in order to reach definitive conclusions. In fact, it is questionable whether qualitative analyses can validly cross-examine large numbers of participants since the expected heterogeneity of the data collected may render the findings highly unsuitable for valid comparisons. In this respect, contrary to quantitative analyses where large participant numbers add validity to a research project, qualitative analyses usually require very small participant numbers since the analyses are so fine-tuned renderings comparisons and conclusions highly debatable. Finally, qualitative analyses are not easily replicable in the sense that individual participants produce differentiated data sets making the whole notion of comparability very difficult to be verified.

Barcroft (2003) investigated the role of input enhancement as to the acquisition of L2 vocabulary “while exploring the role of distinctiveness, the degree
to which an item in the input diverges from the form in which other items in the input are presented, with regard to the nature and direction of the effects of enhancement” (p. 47). Despite the fact that, as Barcroft pointed out, input enhancement directed at vocabulary learning is “qualitatively different from” input enhancement directed at textual comprehension in the oral or written modes (p. 50), there is no sound theoretical reason why Barcroft’s research agenda is not a contribution to the literature on input enhancement.

Within this theoretical framework, Barcroft carried out 2 experiments. The final participant pool in Experiment 1 consisted of 15 1st-semester Spanish L2 learners. They formed 2 groups: Input Enhancement (IE) (9 words out of 24 enhanced) and Unenhanced (unenhanced text, i.e. Control). The final participant pool in Experiment 2 consisted of 21 1st-semester Spanish L2 learners. They formed 2 groups: Input Enhancement (IE) (3 words out of 24 enhanced) and Unenhanced (unenhanced text, i.e. Control). The only enhancement technique Barcroft used was typographical manipulation.

Learning outcomes were assessed through a variety of assessment tests. The combined results of Experiment 1 and Experiment 2 indicated a number of patterns in the data. More specifically, the input enhancement treatment in Experiment 2 (only 3 words out of 24 enhanced) exhibited beneficial learning outcomes as to both the learning of the enhanced words and the learning of unenhanced words, whereas the input enhancement treatment in Experiment 1 (9 words out of 24 enhanced) did not exhibit beneficial learning outcomes in relation to both enhanced and unenhanced words (p. 67 - 68).

Based on these findings, Barcroft suggested that, as to the learning of vocabulary, enhancing about 10% - 15% of the total list of vocabulary items might
prove beneficial (p. 68). This issue of the amount or textual enhancement required to procure the desirable learning outcomes has also been referred to by a number of researchers pointing to tentative detrimental effects on learning due to the over-use of typographical manipulations (Lorch et al., 1995, p. 63; White, 1998, p. 90).

As to the fact that when enhancing a limited number of words, this \textit{birectionality} favourably affects learning of unenhanced words, as well, Barcroft attributed it to the role of “attention and processing resources” which might be affected by the amount of enhancement, i.e. the overall salience of target vocabulary items (p. 69). This statement seems logical if learners’ attentional resources are assumed to be limited as within the Processing Instruction research domain. It also makes sense logically that minimizing the memory load textual enhancements require may enable bigger focal attention on unenhanced material. However, such and similar considerations are meaningful only within small-scale research agendas, or second or foreign language settings with limited amounts of input available and the requirement to procure the desired learning outcomes speedily. These considerations do not apply to naturalistic or naturalistic-like language learning settings where input flooding is \textit{naturally} channelled, and the focal point is acquisition per se and not its rate or speed.

Leow, Egi, Nuevo, and Tsai (2003) investigated “whether type of linguistic item plays a role in noticing and input enhancement” (p. 7). The target structures selected were the Spanish Present Perfect and the Spanish subjunctive. An important consideration in this study was that the authors’ choice of target structures was based on the relative salience of these target structures. According to Leow et al., “the present subjunctive is perceived to be less salient than the present perfect form” (p. 7).

Selecting target structures based on salience has proven both theoretically and practically valid within the input enhancement research paradigm. Theoretically, input
enhancement affects learning through an increase in noticing and awareness which subsequently renders targeted language forms more perceptually salient and easily acquired. Practically, input enhancement has always been a useful tool used by textbook writers (written input enhancement) and teachers (aural – oral input enhancement).

The final participant pool consisted of 72 students attending 1st-year Spanish courses in college. The participants formed 2 groups: Enhanced and Unenhanced. In order to be exposed to both target structures, the subjects were “randomly exposed” to the following conditions:

1. + Enhanced + Subjunctive
2. + Enhanced + Present Perfect
3. – Enhanced + Subjunctive
4. – Enhanced + Present Perfect.

(Adapted from Leow et al., 2003, p. 8)

Learning outcomes were assessed through a multiple choice recognition test and a multiple choice comprehension test. Noticing was assessed through think-aloud protocols and “any written marks related to the targeted verbs” (p. 9). The results of the statistical analyses revealed no statistically significant differences on all 3 measures.

Leow et al. discussed “the failure of textually enhanced input to substantially benefit learners exposed to the enhanced input when compared to those exposed to the unenhanced input” (p. 10, italics added) as tentatively attributable to “learner-external factors” such as the type of tasks employed and the demands of these tasks (p. 10 -
They also pointed out the importance of the perceptual salience of target forms, since the conditions exposed to the Spanish Present Perfect reported more noticing than the conditions exposed to the Spanish subjunctive.

Finally, Leow et al. (2003) correctly underscored 2 important issues. First, the need to incorporate both “perceptive” and productive assessment tasks in L2 research (p. 12). Despite the fact that perception as a variable is very often present in research designs within SLA, the issue of validly assessing perception of target forms remains problematic both theoretically and practically. In theory, perception is an elusive issue especially taking into account cognitive accounts such as gestalt learning (see e.g. Wertheimer, 1959) which postulate that the whole is more than the addition of its parts. The issue is even more problematic in actual research practice since mere performance on a test cannot guarantee the accurate assessment of the exact amount of perception of target forms, especially in the case of very young foreign language learners who lack the necessary linguistic tools to express their perception. Finally, in real life situations the issue is at a quandary since young children implicitly acquire language and are prone to perceive even delicate emotional patterns without showing any signs of such perception. Second, Leow et al. suggested that a “greater amount of exposure time may be needed for learners to notice linguistic forms in a reading passage at a higher level of awareness” (p. 12). The issue of the time needed for noticing targeted linguistic forms has already been addressed previously.

Wong (2003) investigated the learning outcomes of combinations of textual enhancement and simplified input. The target structure selected was the past participle agreement in relative clauses in L2 French. The study also investigated learners’ comprehension of target texts. The final participant pool consisted of 81 English-
speaking students attending 2nd-semester French L2 classes at 2 universities in the USA. A 50% cut-off score was used. The participants formed 4 groups:

1. TE + SI group (textual enhancement + simplified input group)
2. SI only group (simplified input only group)
3. TE only group (textual enhancement only group)
4. Un simplified input without textual enhancement group (comparison group).

(Adapted from Wong, 2003, p. 23)

The textual enhancement techniques used were bolding and italics for targeted forms, as well as underlining for whole clauses. Learning outcomes were assessed through an error correction task (measuring acquisition of target forms) and through free recall tasks (measuring comprehension and recall of total idea units and enhanced idea units).

The results of the statistical analyses revealed the following findings. As to the acquisition of the target structure, even though no differences between the textual enhancement and simplified input groups were reported, the performance of the learners in all groups was better in the post-test than in the pre-test. As to comprehension, for total idea units recalled, only the simplified input only group exhibited enhanced learning outcomes. As to comprehension, for enhanced idea units recalled, only both the textual enhancement groups exhibited enhanced learning outcomes.

Based on these findings, the author proposed a number of explanations. More specifically, Wong argued that textual enhancement does not have an effect on the acquisition of the past participle agreement attributing this non-effect to a host of
factors (p. 31). First, the author attributed it to the non-communicative value of the
target form. Second, Wong speculated that “participants had no incentive to attend to”
the target form, or, in case they had attended to it, this does not guarantee that the
input actually becomes intake as Sharwood Smith had proposed. Finally, Wong
tentatively attributed this non-effect to (a) the underlining that might have distracted
learners from targeted forms and forced them to focus on enhanced idea units, or (b)
the nature of the instructions provided which did not explicitly require learners to
attend to targeted forms.

Wong also argued that simplified input does not have an effect on the
acquisition of the past participle agreement attributing this non-effect to the fact
simplified input may have rendered the texts easier to comprehend but further
attentional resources had been depleted (p. 32). Also, Wong argued that textual
enhancement aided comprehension attributing this effect to the underlining of
enhanced idea units (p. 32 - 33). Furthermore, the author underscored the beneficial
effects of simplified input on comprehension.

Finally, Wong concluded by raising a number of pertinent issues. First, the
need to investigate the relative effects of different textual enhancement techniques (p.
34). Second, the need to investigate the amount of textual enhancement necessary to
procure the desired results (p. 35). The second issue has already been previously
addressed at length. As to the first issue, it is definitely in the right direction since one
of the aims of the input enhancement research agenda is to investigate techniques that
are most optimal for learning. However, this issue sides on individual learning
preferences as well as on the textual enhancement techniques individual learner
populations are more accustomed to. In other words, some individuals may benefit
from a certain technique, whereas other may benefit more from a different technique.
Also, repeated exposure to a certain technique may predispose learners to benefit more through familiarity to the technique. Exact replication with very similar learner populations may tentatively answer the whole issue.

Lee (2007) investigated the relationship between textual enhancement and topic familiarity. The target structure selected was the English passive voice. The theoretical framework of this study was based on a number of issues such as the recent comeback of explicit instruction in SLA within the framework of FonF, the role of textual enhancement “as an optimal intervention point for fulfilling the objective of focus on form”, and the textual enhancement-related research in SLA. Lee pointed out that “textual enhancement is considered more explicit than input flooding but less explicit than rule explanation” citing White (1998) (Lee, 2007, p. 94). The motivation of Lee’s study was to investigate the relationship between textual enhancement and comprehension.

A contribution of Lee’s study was that participants received multiple exposures, i.e. 3 encounters, to the reading material in contrast to most previous single-exposure textual enhancement studies. The rationale was that this would familiarize participants with textual enhancement and that the participants would be facilitated in attentional allocation while processing target forms. This echoes the previous argumentation on learner familiarity with specific techniques. The participants were exposed to various enhanced and unenhanced texts on familiar and unfamiliar topics. Textual enhancement was implemented through “larger, boldfaced letters in different fonts” (p. 97).

The final participant pool consisted of 259 11th-grade high-school South Korean students who had already studied EFL for 4 years. The participants “had previously been taught the target structure, namely the passive voice in English,
through explicit rule presentations by their teachers, but pretest results … showed that they had failed to internalize the target form” (Lee, 2007, p. 96).

This is a serious limitation of Lee’s study and, even though the significance of prior knowledge has already been addressed, some remarks need to be made. More specifically, pre-test results cannot guarantee minimal knowledge of previously taught target forms since (a) the assessment tools may not be sensitive enough to accurately measure such prior knowledge and (b) the assessment tools may exhibit test-taking effects. In Lee’s study, it could be argued that the tests might have operated as a short revision of the target forms without, however, exhibiting signs of learning as to the target forms. Also, Lee’s aforementioned postulation that the learners “had failed to internalize the target form” (p. 96) lacks the specification of the exact meaning of this internalization. Taking into account the relatively young age of these learners coupled with the fact that they had studied English as a foreign language for 4 years further complicates the issue. As to the young age of the learners, these learners might have lacked the necessary linguistic skills to express their knowledge of target forms, especially taking into account that they were assessed through form correction and free-recall tests and their L1 was Korean, a non-Romance language, strikingly different from the target language, i.e. L2 English. As to the learners’ previous learning experience in the L2, i.e. 4 years of study, (a) they had already been taught the passive voice as Lee admits, and (b) they had probably received extensive exposure to related forms such as the verb to be, as well as the English Past Simple Tense which morphologically resembles the past participle. In other words, in all likelihood these learners should have been quite familiar with both constituents of the English passive voice, i.e. the declension of the copula in its various tenses and the past participle. All in all, assuming that Lee accurately reposts lack of internalization
of the target form, this in no way guarantees that these learners had no (implicit) knowledge of (the formation of) the target form.

Lee’s participants formed the following groups:

1. – Enhancement + Topic Familiarity (– E / + F)
2. – Enhancement – Topic Familiarity (– E / – F)
3. + Enhancement + Topic Familiarity (+ E / + F)

(Adapted from Lee, 2007, p. 99)

Learning was assessed through a pre-experimental L2 reading proficiency test, form correction tests, and free-recall tests. The results of the form correction task showed that “the + E / − F group performed best in the form correction task, and the students with enhanced texts (the two + E groups) performed better than those with the baseline version texts (the two − E groups)” (p. 101). The results of the form correction task showed that “(t)he groups given the familiar topic texts (the two + F groups) performed better than the groups that read the less familiar topic texts (the two − F groups). In addition, the students with the baseline texts (the two − E groups) performed better than those with the enhanced texts (the two + E groups)” (p. 103, first parenthesis added).

Based on these findings, Lee argued that textual enhancement assumedly helped learner’s attend to the target form (p. 106). Lee pointed out that these findings were different from other similar findings and attempted to explain the differences along the following lines. First, Lee’s learners were exposed to multiple text encounters. Second, Lee’s learners had been repeatedly taught the target structure
(passive voice) in previous years. In this respect, Lee agreed with Overstreet who had argued that “the implicit nature of the enhancement … was more likely to be beneficial to learners who already had some initial awareness of the forms and their use” (Overstreet, 1998, qtd. in Lee, 2007, p. 106 - 107). Lee also agreed with Overstreet that textual enhancement might exhibit detrimental learning effects on comprehension (p. 107).

As to the findings concerning topic familiarity, Lee pointed out that “enhancement had the effect of making the forms salient to a point where intake was facilitated, but independently from how well the text was understood” (p. 108), indicating a “dissociation between acquisition and comprehension” (p. 108). Lee concluded by recommending textual enhancement combined with explicit instruction may procure superior learning outcomes (p. 110).

Rott (2007) investigated the relative effects of the frequency of input enhancement on vocabulary acquisition and text comprehension. A central point in Rotts’ theoretical framework was the notion of lexical form-meaning connection (FMC) (p. 166). Rott pointed out that even though “text comprehension aims at interpreting the message content, word learning aims at establishing FMCs to build a lexical system” (p. 166). Working memory (WM) is of crucial importance in this respect and Rott adopted the position that probably “WM is a limited-capacity processing system” (p. 166 -167).

As to the assessment measures used within this research paradigm, Rott underscored that receptive and productive tasks have procured contradictory results. Focusing on output-based tasks, Rott cited a number of studies which had reported that “output tasks were less effective for word learning when meaning was not provided explicitly and when readers engaged in output during reading” (p. 168).
Moreover, it had been reported that engagement in output tasks during the process of reading texts disrupted textual comprehension (p. 168). According to Rott, focusing attention on vocabulary items might not be as beneficial as focusing attention on syntactic or morphosyntactic forms. Rott completed her theoretical framework by proposing that “(r)epeated word enhancements increase the potential for establishing FMCs” (p. 171, parenthesis added).

The final participant pool consisted of 38 4th-semester learners of L2 German. The participants were native English speakers. All the participants read 3 texts under 3 conditions: target words were glossed 4 times (4G treatment condition), target words were glossed, “retrieved in the L1”, and bolded 2 times (GR treatment condition), and target words were glossed and “then bolded three times” (GB treatment condition) (p. 171). Learning outcomes were assessed through the following tests: vocabulary pre-test, vocabulary post-tests, and a text comprehension test (oral reproduction). The statistical analyses procured a number of detailed results, the most pertinent of which will be reported at present, along with Rott’s remarks.

Based on the findings of the statistical analyses, Rott stated that when target words are bolded, this does not have significant vocabulary learning outcomes in contrast to “more obtrusive interventions” like supplementary glosses or the retrieval of vocabulary items (p. 186). When readers had both processed the glossed target words and had also come across these words an extra 3 times in a visually enhanced format (GB treatment condition), they could retrieve the targeted words receptively, but had difficulties in actual (re)production of these target words. Nonetheless, “repeatedly directing L2 readers’ attention to semantic word enhancements within the same text has clear advantages for the development of productive word knowledge” (p. 188).
According to Rott, “these findings imply that the visual enhancement of bolding activates qualitatively different attentional resources than semantic enhancements … which resulted in higher levels of word knowledge encoding” (p. 188). Rott admitted an inability to provide a comprehensive account as to the causative variable of the inefficacy of textual enhancement, arguing that possibly the enhancement aided text comprehension but did not assist learners’ focal attention on enhanced words.

Rott’s argumentation is justified since a possible drawback of textual enhancements is that they do not guarantee that (a) the enhanced material will be focused on, and (b) the purpose of the enhancement will be achieved unless explicitly stated. The second postulation indicates that maybe learners should be informed about the purposes of the enhancement. However, this is counter to assumptions that explicitly informing learners about target forms undermines the purpose of implicit or implicit-like treatments.

Finally, Rott explained that increased word frequencies coupled with increased focal attention on semantic meaning may accelerate the speed of learning words productively. It is a fact that increasing the frequency of target words through input flooding is present in many research designs in a disguised manner and sometimes input flooding and textual enhancement as research variables are confounded. Many treatment conditions and assessment measures within SLA research, irrespective of treatment condition employed, include high frequencies of target forms. Taking also into account test-taking effects, this means that input flooding, although not overtly stated as part of a research agenda, is often present, and confounded with various treatment conditions.
Simard (2009) investigated the relative effects of different textual enhancement formats. The target structure selected was the English plural marker. After a review of the research on textual enhancement, Simard reported “contradictory results” (p. 126) and attributed these contradictory findings, among other explanations, to “the number and the type of typographical cues used in the tasks” (p. 126). This observation was the motivation of Simard’s study. In other words, Simard attempted to examine various types of textual enhancement as separate variables in order, among other things, to offer meaningful results that could be comparable with the findings of other related studies.

The final participant pool consisted of 188 “secondary one” French-speaking students attending ESL courses from 2 cities in Quebec, Canada (p. 129). The participants had to read a text (and complete an information transfer task) under 8 different treatment conditions where the English plural markers were enhanced in the following formats:

1. Italic group
2. Underlined group
3. Bold group
4. (yellow) Color group
5. Capital group
6. 5-cues group (italics, underlining, bolding, coloring, and capitalization simultaneously)
7. 3-cues group (bolding, capitalization, and underlining)
8. Control group (no enhancement).

(adapted from Simard, 2009, p. 129)
Learning outcomes were assessed through a multiple choice recognition test. Simard summarized the results of the statistical analyses as follows:

“(A)t the posttest there were significant differences between the Capital and Control groups (F (353) = 2.41, p = 0.02), between the Capital and Underlined groups (F (353) = 1.93, p = 0.05) and between the Capital and 5-cues group (F (353) = 2.40, p = 0.02). The Capital group obtained the best results in all cases. There were also differences between the 3-cue and 5-cues groups (F (353) = _2.14, p = 0.03) and between the 3-cues and Control groups (F (353) = _2.13, p = 0.03). Group 3-cues obtained, in both cases, the best results.

We then grouped all the subjects who were exposed to only one typographical cue and formed a new group: 1-cue. We compared the results of the 1-cue group to the 3-cues, 5-cues and Control groups. No significant difference was observed between the 1-cue group and the other groups.”

(Simard, 2009, p. 131, first parenthesis added, remaining parentheses in the original)

Simard made explicit reference to the fact that capitalization and the 3-cues combination procured the best learning outcomes compared to the rest of the textual enhancement techniques used and continued that no relevant studies had been conducted within the SLA framework. The only comparable studies Simard referred to were related with L1 acquisition.

Simard proposed a combination of the textual enhancement and input processing research domains in order to examine whether target forms of high communicative value are more amenable to such treatment conditions than target forms of low communicative value. Simard’s proposal does make sense from a wider pedagogical perspective. Although conducting research in order to examine the effects of treatment conditions is scientifically verified and justified, the true scope of research is the improvement of learning conditions to the benefit of the learners. In this respect, combinations of treatment conditions are certainly warranted especially if the findings prove beneficial for learning.
Hernández (2011) set out to examine the relative effects of (a) explicit instruction (EI) + input flood (IF), and (b) input flood in isolation. The target structure selected was the Spanish discourse markers to narrate a past event. The final participant pool consisted of 91 university English-speaking 4th-semester students attending Spanish courses at a university in the USA. The participants formed 3 groups: explicit instruction (EI) + input flood (IF) (+ feedback), input flood (IF), and control (C). A “baseline” native speaker group consisting of 4 participants was also included in the research design (p. 164). Learning outcomes were assessed through a picture-description task using digital recorders.

The statistical analyses revealed that as to the frequency of discourse markers despite the fact that both experimental conditions performed the same on the pre-test, they both exhibited significant learning gains in both the immediate post-test and the delayed post-test, compared to the control condition. The statistical analyses also revealed that as to the distribution of the discourse markers both experimental conditions exhibited significant post-experimental gains, even though the EI + IF treatment condition learners used discourse markers more frequently in all 3 speaking tasks. Finally, the statistical analyses of the data transcripts revealed that learners in the IF treatment condition used less and more narrowed in range discourse markers than learners in the EI + IF treatment condition.

Hernández explicitly discussed that adding EI to IF does not procure enhanced learning outcomes compared to IF in isolation. Despite the fact that more studies need to be conducted in order to validly assess Hernández’s postulation, one possible explanation might be that there is both a quantitative and a qualitative dissimilarity between EI and IF. In other words, explicit instruction is based on the premise that learners cannot, on their own, acquire target forms efficiently. In this respect, learners
need to receive explicit information on these target forms in order to aid or speed up acquisition. On the other hand, input flood is based on the premise that learners can and will, in the long-run, acquire target forms on their own. In other words, learners need to receive a high frequency of target forms in the input to aid acquisition. In the case of input flood, the emphasis is not on the rate of acquisition through external means, but on the triggering of the internal learning mechanisms by the learners themselves. The focus of input flood is on quantity and the learner, whereas the focus of explicit instruction is on quality and the instructor. Maybe this can explain why the learners in the EI + IF treatment condition in Hernández’s study failed to outperform the learners in the IF only treatment condition.

As Hernadez concluded,

“the combined effect of EI and IF was not superior to IF in promoting learner use of discourse markers. Rather, results indicate that exposure to a rich IF combined with communicative practice and feedback is sufficient to foster acquisition of discourse markers, despite their lack of salience for second language learners”. (Hernández, 2011, p. 177)

However, Hernández pointed out a serious limitation of the study, namely that since the EI + IF instructional treatment learners also received feedback, it could not be validly determined whether the combination of EI + IF or the feedback included were the causative variable of the enhanced learning outcomes. This is a notable limitation in the research design since the amount of information offered to the EI + IF group of learners was also coupled with negative evidence, something absent from the IF only group of learners. It means that the amount of information offered to both groups was not balanced accurately; consequently, the findings need to be interpreted with great caution. However, theoretically, this is not an issue, since there is, as
aforementioned, a qualitative difference between EI, on the one hand, and IF, on the other hand. More specifically, negative evidence in the form of feedback is an essential part of explicit instruction. On the other hand, in IF the learners are required to master the learning material on their own, without external assistance, thereby invalidating any external assistance such as negative evidence in the form of feedback.

Park and Nassif (2013) investigated the role of input enhancement on the comprehension and production of different target forms. The target structures selected were the Arabic comparative form and the Arabic dual pronoun. An important consideration in Park and Nassif’s study was that their choice of target structures was based on the relative communicative value of the structures. According to Park and Nassif, the Arabic comparative form is characterized as having high communicative value and being non-redundant, whereas the Arabic dual pronoun is characterized as having low communicative value and being structurally redundant. The 2 authors investigated the effects of input enhancement on local comprehension as well as on global comprehension. Local comprehension concerns (the recall) of enhanced information, whereas global comprehension concerns (the recall) of non-enhanced information.

The final participant pool consisted of 16 students, English-speakers, attending an Arabic FL programme spring semester course at a graduate school in the USA. The participants formed 2 groups: Enhanced group (EG) and Unenhanced group (UG). The enhancement techniques Park and Nassif used were font enlargement, bolding, and underlining. Learning outcomes were assessed through 2 comprehension tests (free recall task and comprehension questions task) and 2 production tests (fill-in the-blanks task and sentence production task). The results of the statistical analyses
revealed that as to global comprehension of the target structure, there were no statistically significant differences between the Enhanced group and the Unenhanced group for the comparative form, but as to global comprehension of the dual pronoun form the enhancement of this non-meaningful form affected negatively global text comprehension for the learners. The results of the statistical analyses also revealed that as to local comprehension, there were no statistically significant differences between the Enhanced group and the Unenhanced group for both the comparative form and the dual pronoun form, but as to local comprehension of the dual pronoun the “participants in the UG outperformed the EG group by about one standard deviation unit, which implies a considerable mean difference between the two groups” (p. 10). As to the production of the target structure, there were no statistically significant differences between the EG group and the UG group in controlled production for both target structures, despite “considerable individual variations as illustrated by the large standard deviations” (p. 11).

The issue of standard deviations has received relatively limited attention in the SLA literature, despite the fact that standard deviation values are commonly reported. Large standard deviation values imply non-uniform, varied performances, but in order to offer more definitive conclusions other values of variance need to be reported, such as variance, kurtosis, and skewness. Nonetheless, the significant individual variations reported by Park and Nassif tentatively imply that the treatment conditions did not affect individual learners in a uniform manner. This is expected, to a certain degree, since an implicit treatment like input enhancement, in the absence of external explicit information, leaves individual learners on their own to decipher the underlying patterns of the target forms. Consequently, each and every learner is free to respond to
the input provided in a unique manner, as exhibited by the large standard deviation values procured.

Park and Nassif discussed the results proposing a number of observations. First, that the enhancement of non-meaningful target forms in the input may be distracting for learners negatively affecting comprehension. Also taking into account Overstreet’s (1998) and Lee’s (2007) findings, Park and Nassif (2013) noted that the collective results imply detrimental effects of textual enhancement on the global comprehension of learners and, possibly, visual input enhancement is a distracting variable (Overstreet, 1998; Lee, 2007, qtd. in Park & Nassif, 2013, p. 12). More specifically,

“(t)aken together, the results suggest that different linguistic forms may be differentially susceptible to TE techniques such that enhancing a linguistic form with greater semantic value may facilitate the learners’ local comprehension. On the other hand, enhancing a form with less semantic value may hinder not only the learner’s local comprehension, but also his/her global comprehension.”

(Park & Nassif, 2013: 13, first parenthesis added, italics in the original)

Finally, Park and Nassif underscored the importance of taking into account the learners’ prior knowledge of target forms in input enhancement studies, but they reported inability to reach a definitive conclusion on this issue based on their study.

Winke (2013) investigated the relative effects of input enhancement on the comprehension and learning of grammar. The target structure selected was the English passive form. Treatments were computer-based. Winke’s study was intended to replicate Lee’s (2007) study and examine whether the input enhancement technique procures “trade-off effects” (Winke, 2013, p. 325), i.e. the beneficial effects of enhanced attention to forms has a detrimental effect on attention to meaning.
As to the assessment of learning outcomes and noticing procured by input enhancement, Winke, after examining the various measures used (p. 327 – 329), concluded as follows:

“The methodological quandary is whether failure to measure noticing through first-person accounts necessarily means that learners did not notice … However, … (it) … may be an instance in which researchers are using a test’s sensitivity in measuring a construct to define the construct itself. It is illogical to state that an entity is one thing if you can measure it, another if you cannot, and something different still if individuals can self-report having it. Awareness may have different levels, but it is not the case that noticing itself switches into being a certain level of discovery. In sum, the problem is an issue of veridicality: a potential lack of efficacy with self reports of noticing. Thus, a more objective method of measuring enhancement’s effect on noticing is needed.”

(Winke, 2013, p. 329, parenthesis added)

Winke’s remark is definitely in the right direction. Undoubtedly, it is of great importance in studies with young, beginner L2 learners as participants. Such learners, as well as the majority of beginner L2 learners, lack, as previously argued, the necessary linguistic tools to report noticing of target forms. Consequently, poor results in self-reports may also be interpreted as implying that an assessment measure is not sensitive enough to calculate accurately the amount of noticing. As Winke correctly pointed out, researchers sometimes confound the assessment measure itself with the object of assessment, i.e. the target form, something which leads to erroneous findings and assumptions. This is especially relevant for implicit-like instructional treatments like input enhancement which by definition tap on implicit knowledge sources not easily observable or expressible.

The final participant pool consisted of 55 university English L2 students at a university in the USA, at the intermediate level. The reading text used was an authentic text modified to include as many target structures, i.e. passive forms, as
possible in 2 forms: unenhanced group and enhanced group (target forms underlined and in red). Therefore, it might be tentatively assumed that Winke’s study included input flooding as well in the research design. Learning outcomes were assessed through tests that included sentences including errors of the target form. A free-recall task was also included in the research design in order to measure reading comprehension. The statistical analyses also included assessment of learning and development of the target form, as well as comprehension.

The statistical analyses procured the following results. As to both learning gains and effects on reading comprehension there were no statistically significant differences between the enhanced and unenhanced groups. Winke summarized the results as follows: First, input enhancement, despite drawing learners’ attention to target forms, “did not increase initial processing any more than flooding did” (p. 341). Second, “measurable acquisition” did not occur as hypothesized by other researchers (p. 341).

To explain these learning outcomes, Winke (2013) proposed a number of explanations: First, “not everything that is registered by the senses is … encoded in long-term memory” (p. 341). Second, the lack of task directions might have had a detrimental effect on learning outcomes, negatively affecting motivation. Third, Winke proposed that input enhancement does not procure trade-off effects insisting that it “is a good, implicit pedagogical tool” (p. 342).

Finally, Winke (2013) pointed out that research on input enhancement and input flooding should be more classroom-oriented, since “(f)looding and enhancement are teachers’ tools and, thus, must be studied in the context of how teachers use them” (p. 343, parenthesis added). Winke continued that input enhancement might have beneficial effects only for learners who already have some kind of prior knowledge of
target forms and that explicit information might be a necessary accompaniment for learners who have no knowledge of the target forms selected.

LaBrozzi (2014) investigated the relative effects of various textual enhancement types in a research design using adult learners at a university in the U.S.A. The results revealed a hierarchy of effectiveness of textual enhancement types in order from more beneficial to less beneficial: (a) changes in font size, (b) capitalization, (c) bolding and italics, and (d) changes in font and underlining. It seems that much like the young learners in Simard’s (2009) study, the adult learners responded not so well as to the textual input enhancement (TIE) technique of bolding, i.e. the TIE technique used in the research design of both the pilot study and the main study in the present thesis.

3.3.3. Input enhancement & input flooding: research: summary

The previous section reviewed the research on visual and textual input enhancement, and input flooding from Trahey and White (1993) to LaBrozzi (2014). During the course of these years the pertinent research, despite prolific, has been mainly concerned with the effects of different visual and textual input enhancement techniques on different target groups and for different target structures. This narrow scope is not necessarily a drawback. Viewed under the prism of practical teaching implementations, it means that practitioners can validly implement these teaching methodologies in the classroom. There is a coherent theoretical framework as a point of reference, as well as a battery of techniques to choose from according to the needs of the learners and the properties of the linguistic forms to be taught / learned. In fact, the criticism raised in the previous section was mainly directed to the research designs
of the various studies, the conflation of visual and textual input enhancement with input flooding, and the way control conditions were materialized. In previous paragraphs Winke lamented that research on input enhancement and input flooding is not as classroom-oriented as it should be. I believe that it need not be. The reason is that such techniques, and this especially applies to input flooding, have been utilized for a long time.

3.4. Input Enhancement, Input Flooding & the English Past Simple Tense

3.4.1. Input enhancement, input flooding & the English Past Simple Tense: introduction

The researchers in the studies reviewed in the following section have selected the English Past Simple Tense as a target structure. The participants in these studies are either adolescents or adults. Consequently, there is a notable difference between these studies and the present main study where the participants are young children. Another difference is that these studies have focused on the English Past Simple Tense, whereas in the present main study the focus of investigation is more limited in scope, i.e. the focus is on the regular English Past Simple Tense –ed. Nonetheless, the findings reported can offer valuable information as to the effects of visual and textual input enhancement, and input flooding as to the acquisition of the target structure.

3.4.2. Input enhancement, input flooding & the English Past Simple Tense: research
Sarboland (2012) investigated the relative effects of different textual enhancement techniques on the acquisition of the English Past Simple Tense by Iranian L1 learners. The originality of Sarboland’s work lies on the introduction of a novel textual enhancement technique labelled “choice TE format” which “introduces both correct and incorrect forms in the reading text while the incorrect forms are identified by an asterisk and follow the correct ones so that learners are first exposed to the correct forms” (p. 460). According to the author, this technique is more explicit than the common textual enhancement techniques, since correct and incorrect forms are presented, and, at the same time, not overtly explicit in the absence of any kind of metalinguistic explanation (p. 460). The following sentence is the example Sarboland uses as to the novel technique:

“…The school headmaster (told/ told*) the girl that she (did not like / do not like*) her bad behaviour …”

(Sarboland, 2012, p. 461, italics in the original)

The final participant pool included 156 male learners at the Iran Language Institute learning English. Their proficiency level was pre-intermediate and their age ranged from 14 to 32 years old (mean age 16.33). None of the participants had any kind of prior knowledge of the target structure. The participants formed 5 groups each exposed to the same reading text in which the English Past Simple Tense tokens were enhanced using different techniques: bolding, italics, underlining, choice. The participants in the control condition were exposed to the same text without any type of enhancement. Learning outcomes were assessed through multiple choice recognition pre- and post-tests.
The results of the statistical analyses revealed that the most effective treatments were Underlining and Bolding without any in-between statistically significant differences. Both groups exhibited statistically significant differences in relation to the other 3 groups, i.e. italics, choice, and control. In other words, the participants in the Underlining and Bolding groups outperformed learners in the other groups at a level of statistical significance.

On the basis of these findings, Sarboland proposed a number of explanations. First, underlining or highlighting was the enhancement technique the participants were most familiar with. Second, underlining is a technique that renders enhanced material more salient than other textual enhancement techniques since it involves the addition of a line below the target structure, i.e. an extra feature that attracts more attention in relation to bolding and italics. “In simpler words, underlining is an additive TE format in that the subject sees something extra in addition to the target structure and this is what, we surmise, brings about the noticing of what the researcher means to be noticed” (p. 469, italics in the original). Third, bolding is very similar to highlighting, a strategy these participants were most familiar with.

Finally, in order to explain the inefficacy of the choice format, Sarboland posited 2 explanations. First, the author suggested that participants were confused “as to why one form is asterisk marked as incorrect and the other one is not and is considered correct”, i.e. it distracted learners from the reading task (p. 469). Second, the author proposed that the choice format caused learners to wonder about the reason why a form is enhanced and, consequently, to try and find a rule as to the appropriate use of the targeted form. Very nicely put,

“(t)his brings the form to consciousness and makes the learner alert about the form and this is not what TE is trying to achieve since TE aims at
making the learners acquire the forms while their attention is focused on meaning and the comprehension of the passage.”
(Sarboland, 2012, p. 470, parenthesis added)

Sarboland’s observations are based on a number of sound teaching and learning principles. The most noteworthy is that the efficacy of any (textual enhancement) technique depends, to a great extent, on how familiar learners are with a particular technique. This is a very important issue and Sarboland underscores the individual learners and their preferences, familiarities, and, consequently, their needs as a variable that decisively affects the efficacy of different textual enhancement formats. The issue has received little, if any, attention in the relevant theoretical, research, and in-practice framework in the SLA textual enhancement literature. The importance of the introduction of a novel textual enhancement technique cannot be overemphasized. Sarboland’s choice format is the only original textual enhancement technique in all the studies reviewed in the present thesis. Finally, Sarboland made a very practical recommendation, i.e. that teachers leave learners do the textual enhancement on their own. However, “there is a drawback to this practice and that is the fact that learners, as a result of this practice will grow more conscious of the forms and this is against the purposes of TE and input enhancement in general” (p. 471).

Nahavandi and Mukundan (2013) investigated the relative effects of textual input enhancement on the acquisition of the English Past Simple Tense by elementary level EFL Iranian learners. Nahavandi and Mukundan highlighted the fact that the differences as to the research designs of the various studies within the visual input enhancement framework have rendered comparisons between studies very hard and distinguished between traditional grammar instruction which placed emphasis on the
learners’ output production and textual input enhancement as type of focus on form which aims at altering the way in which learners perceive and process input (p. 93).

The final participant pool included 93 elementary Iranian English L2 learners ranging from 19 to 40 years old. A serious limitation of the participant sample in Nahavandi and Mukundan’s study was “they had learnt the English simple past tense, but had not yet developed full mastery of the form” (p. 94). This means that they had not only been exposed to the target form, but they had in fact learned it. Coupled with the fact that they had not fully mastered the target form implies that these learners had more than adequate prior knowledge of the past simple tense. The very high 90% cut-off score used raises further doubts about the actual effectiveness of the treatment conditions. All these details necessitate a cautious interpretation of both the findings and the conclusions drawn from the findings. The participants formed 3 groups: Textual Input Enhancement only, Textual Input Enhancement + rule presentation, and Control. The instructional treatments included 3 passages including multiple choice comprehension tasks. The enhancement technique used was the bolding of English Past Simple Tense target forms. The control group were exposed to the same passages without bolding. This suggests that it was not a true control group since it received exposure to the target form probably through input flooding. This further necessitates a very cautious interpretation of both the findings and the conclusions drawn from the findings. Learning outcomes were assessed through multiple choice recognition tests in a pre-test, post-test, and delayed post-test design.

The results of the statistical analyses revealed a significant difference among the 3 instructional treatment conditions with the Textual Enhancement + Rule Presentation group being the only one to statistically outperform the Control condition. Nahavandi and Mukundan also analyzed statistically the existence of
statistically significant differences based on the gender of the participants. However, no statistically significant differences were reported.

Nahavandi and Mukundan drew a number of conclusions based on their findings (p. 99 -100). First, that both the textual enhancement and the textual enhancement + rule presentation treatment conditions positively aided learners’ intake of the English Past Simple Tense. Second, that the combination of textual enhancement and explicit instruction in the form of rule presentations enhances the learning outcomes procured by textual enhancement alone. This final remark seems to imply that combinations of implicit (textual enhancement) and explicit (rule presentation) instruction treatment condition may prove beneficial as to the acquisition of the English Past Simple Tense. Despite the fact that further implications could be posited, the fact that the control condition was also exposed to the target structure, coupled with the fact that participants did have prior knowledge of the target structure, does not validate further discussion of Nahavandi and Mukundan’s argumentation. However, the basic finding that textual enhancement coupled with explicit instruction proved beneficial as to the acquisition of the English Past Simple Tense is a self-evident fact.

Rikhtegar and Gholami (2015) investigated the effects of the timing of input flooding, i.e. presentation of the target structure before and after exposure to texts which included input flooding to the targets structure. The target structure selected was the English Past Simple Tense. The theoretical framework of Rikhtegar and Gholami was placed within input-based instruction and the various forms of input enhancement that can be used to aid instruction on different aspects of grammar. According to the authors, “grammar instruction is an imperative matter to be considered in Iranian classes” (p. 80).
The final participant pool included 61 mostly Turkish L1 elementary proficiency level learners aged between 11 to 14 years old. Through a variety of proficiency tests, a 60% cut-off score was used. The participants formed 3 groups: G1, G2, and Control. G1 were exposed to a reading passage including an input flood on the English Past Simple Tense prior to rule presentation on the target form. The rule presentation included a combination of deductive and inductive techniques. G2 were exposed to a reading passage including an input flood on the English Past Simple Tense after rule presentation on the target form. G3 (control) were a true control condition in that they simply followed their regular school curriculum without any kind of exposure to the aforementioned techniques. They just attended “the normal process of presenting the grammar” in their school (p. 83). Rikhtegar and Gholami did not explicitly mention whether this normal process included exposure to the target form or not for the control condition. Learning outcomes were assessed through the administration of a pre-test and a post-test prepared by the teachers, as well as through a dictogloss task.

The statistical analyses conducted revealed the following results: First, at to the acquisition of the English Past Simple Tense, both the G1 and G2 groups outperformed the control condition at a level of statistical significance. There was no statistically significant difference between G1 and G2. Second, as to the production English Past Simple Tense assessed through the number of errors produced by the 3 groups, both the G1 and G2 groups outperformed the control condition at a level of statistical significance. This means that both treatment conditions made fewer errors on the target form than the control condition. However, there was a statistically significant difference between G1 and G2. More, specifically, the post-presentation
input flooding group (G2) outperformed the pre-presentation input flooding group (G1) in this respect.

On the basis of these findings, Rikhtegar and Gholami argued that input flooding including rule presentation both before and after reading beneficially affected both the acquisition and the production of the English Past Simple Tense to a similar extent. However, input flooding including rule presentation before reading procured even more beneficial learning outcomes on accuracy of production compared with input flooding including rule presentation after reading. The 2 “researcher teachers” concluded by a call for including input flooding combined with the presentation of rules in both research and teacher training courses (p. 87).

3.4.3. Input enhancement, input flooding & the English Past Simple Tense: summary

The 3 studies reviewed in the previous section, i.e. Sarboland (2012), Nahavandi and Mukundan (2013), and Rightegar and Gholami (2015) investigated the acquisition of the English Past Simple Tense (both regular and irregular) through the use of textual input enhancement techniques. It is impossible to reach definitive conclusions due to the fact that these 3 studies utilized very different research designs, a common problem of research in this field of science as Nahavandi and Mukundan (2013) pointed out (p. 93). A notable commonality of these 3 studies is that they were carried out within the context of the educational system of Iran. Another similarity among the 3 studies is that the researchers seemed preoccupied with the practical implementation of their findings in actual classroom practice. In all likelihood the choice of target structure reflects this preoccupation. In contrast to the rest of the studies within the textual and visual input enhancement, and input flood framework
reviewed in section 3.3, these 3 studies were more-classroom oriented, reflecting the priorities of their educational context.

The cumulative findings of these studies can not offer definitive conclusions as to the efficacy of textual and visual input enhancement, and input flood on the acquisition of the English Past Simple Tense. Nonetheless, they are valid proof that textual and visual input enhancement, and input flood can and should be researched in the context of the acquisition of this particular target structure.

3.5. Processing Instruction: Research

3.5.1. Processing instruction: research: introduction

In this section research within the framework of Processing Instruction (PI) will be reviewed. This presentation does not include studies incorporating Structured Input (SI) in their research design. These studies will be presented in a subsequent section. Research on processing instruction began with the VanPatten and Cadierno (1993) study and remains prolific. Despite differences, these studies include a number of common elements, the most notable of which is that PI (an input-based type of instruction) is usually compared with various types of output-based instruction, e.g. TI, MOI, MOBI, etc. Another commonality concerns the assessment measures used. These measures most of the times include various types of interpretation / comprehension and production tests. It seems that one of the major concerns of research in this paradigm is the inter-relationship between input and output. Furthermore, almost all of the studies within the PI research paradigm include an explicitly-stated attempt to alter learners’ erroneous processing strategies with the
ultimate aim to foster the corresponding correct processing strategies. In fact, despite the fact that recently PI studies have broadened their horizons and scope, the issue of processing strategies is the most notable common feature of such studies.

In this respect, it seems that both PI theory and research constitute a research framework in which the learner is hindered by language-specific erroneous strategies that may be an attribute of the L1 or the L2. Learners are unable to surmount these processing strategies on their own. Consequently, it is the task of the researcher / instructor to appropriately inform learners about the erroneous strategies and present the correct ones. After such brief presentations, learners are supplied with input that has been meticulously structured to help learners practice / process the correct strategies. In PI theory and research the focal point is input processing and not output production and, despite the fact that the role of output is not ignored, learners are not usually required to practice producing output. In other words, PI is exclusively input-based. However, a standard concern of researchers is whether such exclusively input-based instruction can procure enhanced learning outcomes on output-based production measures, as well.

A basic misconception of both the proponents and the adversaries of PI-centered research has to do with the assumption that input-based PI should naturally load on interpretation measures by definition and output-based types of instruction should procure enhanced learning outcomes on production measures de facto. PI advocates are usually concerned with the beneficial effects of PI instructional treatments on production measures, as well, whereas, PI critics usually emphasize enhanced learning outcomes on production measures by output-based instructional treatments.
Both lines of reasoning, however, downplay the most fundamental factor of learning, i.e. learning itself, and, instead focus on the measurable outcomes of learning. To be more specific, any type of learning requires some kind of input to trigger the learning process. The ability or inability of a learner to perform on (interpretation or production) assessment measures is irrelevant to learning. Instead, such measures do not really measure learning per se, but the materials designer’s perceptions of learning. In other words, both advocates and critics of PI research erroneously assume that enhanced performance on whatever type of assessment measure they utilize is synonymous with learning / acquisition. In fact, it is nothing else but a sign of learning. I would like to emphasize that I do not believe that such assessment measures are totally inefficient. My point is that instead of focusing on whether input-based or output-based instruction are the most optimal learning tools, it is high time we reviewed our assessment measures.

Despite the fact that learning-centered research has recently broadened its battery of assessment tools, the fundamental misconception remains. The misconception that what counts is the output / outcome of instruction and not learning itself. In this respect, research within the PI research framework is not an exception.

3.5.2. Processing instruction: research: studies

VanPatten and Cadierno (1993) investigated the relative effects of PI and TI using as a target structure object pronouns in Spanish. The strategy the researchers sought to alter was the First Noun Strategy. At the beginning of the study, VanPatten and Cadierno attempted to examine the following research questions:
1. whether learners’ developing system (IL) is affected by the way learners process input,
2. in case the developing system (IL) is affected, if the effect is transferred to output or it is limited to input,
3. in case the developing system (IL) is affected, if the effects of PI are the same as those of TI.

(adapted from VanPatten and Cadierno, 1993, p. 229)

The 2 types of instruction were operationalized differently. TI involved presentation through a paradigmatic chart, grammar explanations, and descriptions with examples, along with information on the placement of object pronouns within the sentence. As far as practice was concerned, learners were offered a number of activities ranging from mechanical form-oriented, to meaningful and, finally, to more open-ended communicative practice. Production was required in both the oral and written modes.

On the other hand, for the PI group the presentation phase began by first examining the contrast between subject and object and then by presenting subject and object pronouns. Explanations regarding the position of the target structure in sentences followed. Subjects were informed about the first noun strategy and some other details on related forms. Learners practiced through 2 types of SI activities, i.e. referential and affective, in the order presented here. A C group continued with their regular schedule without any kind of instruction on object pronouns.

The final participant pool consisted of 80 2nd-year university students learning Spanish in a university in the USA. According to VanPatten and Cadierno, “subjects were not scheduled to receive explicit instruction in object pronouns and word order.
during the time period of … (the) … investigation” (p. 229, parenthesis added). These participants formed 3 groups: PI, TI, and C. Learning outcomes were assessed through an interpretation, a distractor, and a production task. For the interpretation task, subjects had to match a sentence they heard with 1 of 2 pictures presented to them on OHP. The distractor consisted of answering questions unrelated to the target structure. In the production task, subjects were required to complete incomplete sentences according to visual clues. An 80% cut-off score was used in both assessment measures. The results of these tests were submitted to one-way (ANOVA) analyses. Accordingly, the PI group performed significantly better than the TI group and the C group in the interpretation test. In the production test, the differences between the PI group and the TI group were statistically insignificant and both improved in relation to the C group.

On the basis of these results, VanPatten and Cadierno answered the research questions claiming that (a) yes, the developing system is affected by the way learners process input (results on the post-experimental interpretation task), (b) this effect is not only limited to input but is evident in output (PI performing as high as TI on the post-experimental production task), and (c) PI is overall superior to TI (combined results) (p. 236 – 237). An important point worth noting in the discussion of the results is that VanPatten and Cadierno did not propose abandoning communicative activities focused on output. To the contrary, they simply suggested that these activities follow input-focused ones.

Cadierno’s (1995) study investigating the Spanish Past Tense was carried out about 2 decades before the present main study which also examines the acquisition of the Past Simple Tense, this time in English by Greek L1 learners. The main study research design shares many commonalities and has borrowed extensively from this
and the original VanPatten and Cadierno (1993) study, as well as Allen’s (2000) study. Nonetheless, both the theoretical underpinnings and the results are different. The same applies to the general theoretical and practical framework derived from the differentiated results.

Cadierno (1995) investigated a beginner level L2 learning strategy: the so-called *Lexical Strategy*, “according to which L2 learners prefer processing lexical items to processing grammatical information” (p. 182). Learning outcomes were assessed through an interpretation and a production task. The results of the statistical analyses showed that the PI group improved significantly in both tasks, whereas the TI group improved only in production. The PI group and the TI group did not generate significantly different results in the production task according to the statistical analysis. The C group showed no signs of improvement.

Cadierno viewed these results as signifying that PI had altered the way subjects processed input, something that consequently affected their developing system (interlanguage) and their abilities in production. Cadierno emphasized that, theoretically, PI does have an impact on second language acquisition and that methodologically IP is an input-based approach close to input-based communicative approaches.

Citing Cadierno and Glass (1991), Glass and Cadierno (1990), and Musumeci (1989) (qtd. in Cadierno, 1995, p. 182), Cadierno postulated that “when learners are confronted with utterances in which both adverbs and verbal morphology are present, they tend to rely on adverbs for temporal distinctions” (p. 182).

“In other words, learners prefer processing lexical items (adverbials) before morphological items (past tense verb morphemes) in tense assignment.”

(Cadierno, 1995, p. 182, parentheses in the original)
Although the Spanish and the English Past (Simple) Tenses differ, the present main study shares a common goal with Cadierno (1995): to encourage “learners to pay attention to non-salient grammatical verbal forms in the input”, i.e. “to investigate the impact of attempting to alter learners’ processing of the past tense” (p. 182).

The present main study, however, is dissimilar to Cadierno (1995) in 2 more than fundamental interrelated ways: (a) the age of the subjects, and (b) the proficiency level of the subjects. As to the age, the 8-year-old learners in the present main study (a) had a much narrower past time conceptualization, (b) mastered a much more limited vocabulary size, (c) had never been explicitly instructed on this target structure, and (d) were naturally less susceptible to explicit grammar explanations and instruction that makes use of terminology like morpheme, Past Simple Tense, time adverbials, and so on; all this compared with the 3rd-semester university students in Cadierno’s study who were probably more highly motivated and accustomed to a context-rich diet of (explicit) instruction on the whole. I consider (a), i.e. the subjects’ scope of past time conceptualization, an unbridgeable difference between the 2 studies. All Cadierno’s subjects had to do was a knowledge schema transfer from talking about the past in English to talking about the past in Spanish. The young children in the present main study had to tackle with the notion of pastness, with limited vocabulary, and a fraction of learning experience.

The second sharp contrast, related to the first if one reads between the lines, deals with the cut-off score used in the Cadierno (1995) study and presents a sharp methodological disparity between the present main study and Cadierno’s study. It is difficult to come up with any sound methodological reason behind arbitrarily using cut-off scores, other than the wish to ensure that subjects have little knowledge of a
target structure. Such arbitrary cut-off scores cannot be sustained methodologically, especially if one takes into account 2 notions of elementary statistics: the Poisson Law and the Gaussian Bell Curve, a theme which is to be discussed in the Discussion section of the present thesis.

VanPatten and Sanz (1995) were motivated by an inadequacy of previous studies concerning the type of assessment tasks, i.e. that these did not test communicative performance. Their aim, in other words, was to examine whether the positive effects of PI were evident in a wider variety of tasks including spontaneous language use. The target structure was again object pronouns and word order in Spanish. They used only 2 groups: (a) PI and (b) NI (i.e. no instruction, control group). Learning outcomes were assessed through the following tasks: (a) the same tests as in VanPatten and Cadierno (1993), (b) a question and answer interview, and (c) a video narration task. Possible effects were sought in both the oral and written modes. The findings were in support of PI: the PI group improved in both speaking and writing and, more importantly, it had an effect regardless of test type, with the exception of the video narration task.

VanPatten and Sanz acknowledged that scores differed according to mode and task type but the results supported their hypotheses. What is more, they noted that the use of multiple assessment tasks is both informative for teachers interested in what works best and researchers examining the full impact of instruction. Still, their results provided no definite conclusions on the durability of instruction. Moreover, the study did not include an output-focused instructional treatment, i.e. TI, so the possible effects of this type of treatment on such tasks cannot be assessed.

DeKeyser and Sokalski (1996) attempted to replicate the original VanPatten and Cadierno (1993) study and test its generalizability. They attempted to examine the
acquisition of both direct object clitic pronouns and conditionals in Spanish. Three groups were formed: (a) input practice, (b) output practice, and (c) C, control.

The results were different for each structure: input practice was more advantageous for the direct object clitic pronouns, whereas output practice was more beneficial for the conditionals. This means that the effectiveness of instruction depended on the given target structure. Moreover, DeKeyser and Sokalski also reported that after analyses they discovered that output practice was better for production and input practice was better for comprehension as predicted by skill acquisition theory. They argued that the morphosyntactic complexity of the target structures coupled with the delay between practice and testing were the causative variable that affected the relative effectiveness of comprehension and production practice.

Based on these findings, DeKeyser and Sokalski (1996) criticized VanPatten and Cadierno (1993) and Cadierno (1995) on 3 issues: (a) the PI group and the TI group had not been offered the same amount of declarative knowledge, (b) despite the fact that the PI group activities had required attention to both meaning and form, the TI group activities had focused on form exclusively, and (c) the choice of structure had affected the results. This final remark was directed only to the VanPatten and Cadierno (1993) study. As a confirmation of this last point they cited Nagata (1995) who had investigated the Japanese honorific system and had reported the exact opposite results from VanPatten and Cadierno (1993). In Nagata’s (1995) study, the output practice group outperformed the input practice group in the production tests and performed almost equally well in comprehension tests (Nagata, 1995, p. 19 - 20, qtd. in DeKeyser and Sokalski, 1996, p. 620).
These 3 points were probably in the right direction. To begin with, a closer inspection of the VanPatten and Cadierno (1993) instructional materials does reveal that there is quantitative, as well as a qualitative, difference in the amount of knowledge offered. In addition, TI cannot easily be equated with attention to form; the issue of meaning has been of paramount importance in a number of traditional (by now) approaches. However, VanPatten and Cadierno were in all probability justified in focusing on form in TI since this was the standard practice in the majority of coursebooks. As for the third issue, it still remains unanswered.

As previously mentioned, VanPatten and Oikkenon (1996) concluded that SI is the causative variable responsible for the learning outcomes in that and previous similar studies. This study is reviewed in a following section of the present thesis presenting studies including SI in their research design.

Salaberry (1997) investigated the relative effects of input processing instruction (IP) and output processing instruction (OP). The target structure selected was Spanish clitic pronouns. The final participant pool consisted of 65 university students attending an L2 Spanish course in a university in the USA. The participants formed 3 groups: PI, OP, and C. Salaberry reported no “significant differences among groups with respect to their degree of exposure to the target language in school or family settings or travel in Spanish-speaking countries”, according to a background questionnaire (p. 428).

Learning outcomes were assessed through 3 tasks: (a) a comprehension test, (b) a production test, and (c) a written narrative on a 1-minute silent video. The statistical analyses revealed that in the production test there were no statistically significant between-group differences. The same applied to the written narrative. In the comprehension test, both input-based and output-based groups performed
similarly and higher than the control group with effects lasting after one month. These results led Salaberry to argue that the benefits of IP can also be guaranteed by Swain’s *Pushed Output Hypothesis*, as well. Most importantly, he believed that both IP and OP can result in the formation of form-meaning connections.

Salaberry claimed that IP is an incomplete account of L2 acquisition, “a simple behaviouristic model”, unable “to subsume all the inherent complexity of the process of SLA” (p. 433 - 434). These seem like rather strong accusations despite the fact that some of Salaberry’s claims might have been on the right track. His line of reasoning was lengthy and can be summarized for the present purposes as follows:

1. VanPatten and Cadierno’s (1993, p. 238) contention that IP leads to acquisition, whereas TI (OP) leads to learned linguistic knowledge is theoretically unwarranted. Moreover, the theoretical division between IP and OP has not been clarified
2. The distinction of the cognitive processes generated by IP (acquisition of the L2 system) and OP (access to the L2 system) appears unfalsifiable
3. IP is not a learner-centered but a teacher-centered approach
4. The role of comprehensible input is exaggerated
5. The identification of linguistic forms to be taught remains a problem
6. The forms chosen may also be learned through explicit knowledge.

(adapted from Salaberry, 1997, p. 424 - 426)

It is plausible that the strongest claims are 1 and 2. Nowhere in the results of VanPatten and Cadierno is the equation OP equals learned knowledge evident although the two researchers implied this (VanPatten and Cadierno, 1993, p. 238).
The inference that IP equals acquisition cannot be substantiated either. In fact, as Salaberry (1997) rightly noted the learning / acquisition distinction is indeed unverifiable. Although it may look like a good idea intuitively, there is no way of knowing whether a piece of knowledge is acquired or learned. In teaching practice, the 2 terms are synonymous. Just because a grammatical form is taught through TI / output practice is not a necessary and sufficient condition that it is learned. On the same basis, a grammatical form presented through IP is not necessarily acquired.

The arguments provided in 3 - 6 are differentially valid. As for 3, IP is probably both learner- and teacher-centered. It is learner-centered in that it takes into account learners’ processing mechanisms and teacher-centered in that it sets out to alter them by manipulating the input accordingly. Furthermore, the role of comprehensible input is to a certain extent overemphasized (point 4) but it must be remarked that VanPatten and associates did recognize a role for output practice (e.g. Lee and VanPatten, 1995, p. 116 - 132, chapter 6). What is more, the forms that are possible candidates for PI (point 5) had indeed not been adequately studied with the exception of very few structures mainly in Spanish. VanPatten’s proposal that processing strategies ought to be altered limited (and still limits) the field substantially. Statement 6 is self-evident; forms can and are taught through explicit instruction with significant success (Norris and Ortega, 2001, p. 178).

Salaberry (1998) also made a number of interesting observations. First, he proposed that the term processing instruction was misleading and not capable of capturing the essential differences between input-and output-based activities (p. 275). Secondly, and this was certainly justified, he attacked Sanz and VanPatten (1998), and, consequently, VanPatten and associates, for generalizing from studies to pedagogical applications but at the same time rejecting modifications to their
instructional materials to test their claims (Salaberry, 1998, p. 276). This is evident from the detailed criticism that VanPatten (2002) made to all the studies that had reported contradictory results for PI. It is possible that in some cases (e.g. Allen, 2000) the modifications were such that neither PI nor TI were biased compared to VanPatten and Cadierno (1993). Thirdly, Salaberry pointed to an important and serious inconsistency in the cut-off scores that eliminated subjects across studies (e.g. VanPatten & Cadierno, 1993: 80%; Sanz, 1994: 60%, reference provided by Salaberry (1998); VanPatten & Oikkenon, 1997: 70%).

On the other hand, Sanz and VanPatten (1998) criticized Salaberry (1997) for not taking into account the oral vs. written mode differentiation and that his 0, 1 scoring system could not reveal potential intermediate effects of instruction (Sanz & VanPatten, 1998, p. 265, 266).

Lee, Cadierno, Glass, and VanPatten (1997) investigated the Lexical Strategy. As a target structure, they focused on learners’ ability to make tense assignments. The final participant pool consisted of 102 university students attending Spanish classes and belonging to 3 different semesters: 1st semester, 3rd semester, and 5th semester. The learners of each level were grouped in 2 conditions that could be described as: (a) “grammatical + lexical cues”, and (b) “grammatical cues only” (p. 8, italics in the original). Learning outcomes were assessed through a free recall / reconstruction of content test and a tense identification test. In the former, subjects who had received lexical cues in the input performed 3 times better. In the latter, no statistically significant differences were reported. Subjects in both tests showed improvement related to their proficiency level, i.e. the higher the level, the better they performed. Lee et al. concluded that learners with less than 60 hours of training are incapable of detecting temporal cues in the input, whether lexical or grammatical (p. 16).
Collentine (1998) investigated the relative effects of PI and an output-based type of instruction (OI). The target structure selected was the Spanish subjunctive in adjectival clauses. Collentine abstained from focusing on a specific IP processing strategy (p. 579). The final participant pool consisted of 54 university students attending a 2nd-semester Spanish course in a university in the USA. The participants formed 3 groups: PI, OI, and C. A series of procedures was used to verify that there were no statistically significant differences between groups. Learning outcomes were assessed through interpretation and production tasks. The results of the statistical analyses were as follows: In both tasks both treatment groups outperformed the control group, and in both tasks both groups performed equally well.

Collentine attributed these differentiated results to the relative communicative value of the subjunctive. More specifically, in her study, Collentine had raised the communicative value of the subjunctive for both the PI and the OI groups (p. 585). The main factor that distinguished the 2 experimental groups was only processing mode, i.e. input or output. These results were interpreted as indicating that the effectiveness of PI was to be found in the raising of communicative value rather than input mode. Another observation is that, as Collentine noted, it is probable that output can also generate changes in learners’ interlanguage systems (p. 585).

Nagata (1998) also reported findings different from VanPatten and Cadierno (1993). His experiment involved the development of 2 computer programs: (a) an input-based program (explicit instruction plus comprehension exercises), (b) an output-based program (the same explicit instruction plus production exercises). The target structure selected was Japanese honorifics. The final participant pool consisted of 14 college-level students attending a 2nd-semester Japanese course. Learning outcomes were assessed through comprehension and production tasks. The statistical
analyses revealed that the output-focused group performed significantly better than the input-focused group in the production tests and equally well in the interpretation tests.

Although Nagata did not make a comparison with the VanPatten and Cadierno (1993) results, his findings were in contrast with the conclusion that instruction when focused on input is more beneficial than when focused on output (VanPatten & Cadierno, 1993, p. 240). In the theoretical discussion on input-based and output-based instruction, Nagata agreed with Swain who had emphasized the role of output and accepted DeKeyser and Sokalski’s (1996) observation that effectiveness of instruction depends on the complexity of the target structure. He added that this effectiveness also depends on the types of tasks learners are required to perform (Nagata, 1998, p. 34).

Since based on computer-assisted learning, Nagata’s findings are not easily comparable with VanPatten and associates’ studies and Nagata abstained from making such a link. However, they are indeed evidence that the role and the effectiveness of output needs to be examined more carefully, especially due to the fact that, as this researcher has shown, it might prove more beneficial than input.

Allen (2000) investigated the relative effectiveness of PI and TI. The target structure selected was the French causative. The relevant IP principle was the First-Noun Strategy. The final participant pool consisted of 179 high school students attending 4th-semester French classes in 3 high schools in the USA. “The students had not previously been introduced to the causative construction, but they had studied the verb faire (“to do” or “to make”), which is used in the French causative” (p. 74, parenthesis and quotation marks in the original). An arbitrary 60% cut-off score was used. The participants formed 3 groups: TI, PI, and C. Learning outcomes were
assessed through interpretation and production tasks. It must be noted that in this experiment Allen pointed out that she had followed VanPatten and Cadierno (1993) as faithfully as possible in all parts of the study (Allen, 2000, p. 72). The results of the statistical analyses, however, were more in accordance with Nagata (1998) than with VanPatten and Cadierno (1993). More specifically, on the interpretation task both instructional treatments were equally effective, while in the production task subjects in the TI group outperformed those in the PI group. These findings were in sharp contrast with VanPatten and Wong (2004) who reported the exact opposite pattern for the same target structure and the same language.

Allen herself noted that the interpretation task in her study and VanPatten and Cadierno (1993) were “very similar” (Allen, 2000, p. 76). She acknowledged “that the production tasks in both studies differed” (p. 79). Perhaps Allen’s study imparted some of the most convincing evidence that questioned the role of PI compared with TI. This was due to the conspiracy of 2 factors. The first was that she had replicated faithfully the original study and the second was that her results confirmed other studies as well.

Allen criticized DeKeyser and Sokalski (1996) and Salaberry (1997) on the grounds that: (a) they provided the same explicit grammar instruction to both the PI and TI groups, (b) the operationalization of PI was different from VanPatten and Cadierno (1993), and (c) no processing strategy was altered (Allen, 2000, p. 72).

Benati (2001) investigated the effects of PI using as a target structure the Italian future tense. The relevant processing strategy was the Lexical Strategy. He compared a PI group and an output-based instruction group and there was a C group, as well. The final participant pool consisted of 39 university 2nd-semester students of Italian in a university in the UK. Among the criteria for inclusion in the participant
pool, no “previous knowledge or linguistic knowledge of the target language” was required (p. 104). Learning outcomes were assessed through 3 tasks: an aural interpretation test and 2 production tests: a written completion text and an oral limited response task. Benati reported that for the interpretation task PI was better than the other 2 groups but the output-based group also improved from pre-tests to post-tests. Combining the production tasks, both the PI group and the output-based instructional treatment group improved but there were no statistically significant differences between the 2 groups.

The reason Benati offered for the improvement of the output-based group on the interpretation task was that the activities for this group were more “meaning-oriented” (p. 116). He claimed that “it is possible to argue that subjects in this group have created some meaningful output and this has served as input to the others in the class” (p. 116). This is also criticism to other related studies as Benati implied that they were biased against TI by creating less meaningful-oriented activities than PI which was obviously predominantly meaning-focused (p. 116). The results in the production tasks were attributed to the beneficial effects PI concurs on the derivation of intake from input. Benati reported durability of effects 3 weeks after instruction.

Benati, as already mentioned, pointed to a limitation in the Cadierno (1995) study. Benati’s objection was that the activities in the PI instructional packet were mostly communicative and focused on meaning while the TI activities were mostly mechanical and focused on form. This is an issue of definitions depending on how one defines PI and TI. The problem does not concern PI but is evident in TI. There is no conclusive evidence to support the view that TI is de facto mechanical and form-oriented. If this stance is taken, i.e. the one Benati (2001) criticizes Cadierno (1995)
for taking, it seems like a bias against TI and in favor of PI, a bias that might have affected the results.

Cheng (2002) investigated the relative effects of PI and TI. The target structures selected were _estar_ and _ser_ (two verbs in Spanish) with past participles and adjectives. The choice of structures was made based on their low communicative value and their semantic details. This renders these verbs difficult for foreign language learners of Spanish. More specifically, PI “in this study aims at altering beginning learners’ strategy of using _ser_ as a default copula” (p. 310). The final participant pool consisted of 109 college students attending a 4th-semester course in Spanish. A 60% cut-off score in the pretests was used. The participants formed 3 groups: PI, TI, and C. Learning outcomes were assessed through 3 tasks: (a) a sentence interpretation task, (b) a sentence production task, and (c) a guided composition task (production). The preliminary statistical analyses for all 3 tasks revealed almost similar superior effects for both the PI group and the TI group in comparison to the C group. However, when the data where limited to the verb _estar_ only, in the interpretation task the PI group was more effective than the TI group that did not differ from the control treatment. In the other 2 production tasks, the PI group and the TI group performed equally well.

Cheng attributed the performance of the TI group (at least in the combined results) to the inclusion of more meaningful and communicative activities than VanPatten and Cadierno’s (1993) equivalent group (Cheng, 2002, p. 316). Furthermore, he argued that the (combined) results may have been attributable to the low communicative value of the grammatical items _ser_ and _estar_ something which might render these forms less susceptible to instruction than e.g. clitic pronouns (p. 316, see also VanPatten, 1987, p. 67).
Cheng (2002) summarized the inadequacies of previous studies that reported results different from the original VanPatten and Cadierno (1993) study. These remarks are provided although they reiterate VanPatten’s (2002) reservations which will be mentioned promptly, as they are concise and well expressed. More specifically, Cheng (2002) noted the following problems in those studies:

1. “Different research designs from the original study when comparing input comprehension vs. output production instructions.
2. Lack of theoretical support of the processing strategies of target form or structure.
3. Inadequate design of the claimed processing instruction.
5. The difference between traditional instruction and output-based instruction that focused more on meaningful and communicative activities.” (Cheng, 2002, p. 309)

Cheng implied that IP may not be beneficial for all kinds of forms. He noted that the higher the communicative value of a form, the more it may make it susceptible to instruction of different types (p. 316). He also raised the issue whether the effects of instruction were durable (p. 317, see also Benati, 2001, p.119; VanPatten & Sanz, 1995, p. 184).

DeKeyser, Salaberry, Robinson, and Harrington (2002) apart from pointing to some inadequacies of studies reporting beneficial PI findings, pointed to 2 factors - one dealing with theory and one with methodology. On a theoretical level, they reminded that VanPatten’s contention that attention is a limited processing capacity had not been in line with more recent attentional theory that assumed that the capacity to attend to input is unlimited (p. 807). Since attention is a central concept in IP, the implication that it is dated is a serious accusation. On the level of methodology, DeKeyser et al. repeated arguments that in VanPatten and Cadierno (1993) the PI
group was supplied with more explicit information than the TI group: PI subjects had more than necessary explicit information for the production test, TI subjects had less than necessary explicit information for the interpretation task (DeKeyser et al., 2002, p. 814).

VanPatten (2002) attempted to provide explanations for the differentiated outcomes born out by the studies that had been carried out thus far. The thrust of his argumentation lied on the fact that he did not believe that a number of studies were conceptual replications of VanPatten and associates’ studies (referring to DeKeyser & Sokalski, 1996; Salaberry, 1998; Collentine, 1998; and Allen, 2000), although Collentine (1998) did not claim to be replicating a study. For DeKeyser and Sokalski (1996) he noted that the hypotheses and the research questions (between DeKeyser & Sokalski, 1996 and VanPatten & Cadierno, 1993) were not the same. In this respect, DeKeyser and Sokalski (1996) had, according to VanPatten (2002), not attempted to alter learners’ processing strategies and instructional groups were not PI and TI but input and output practice respectively. The fact that his study had not attempted to alter learners’ processing strategies was also directed towards Salaberry (1997). As far as Allen (2000) was concerned, VanPatten (2002) posited that although her study had seemed like a replication on the surface, it had fallen short in 3 crucial points: (a) the TI instructional materials were not pure TI but a blend of PI and TI, (b) PI did not make the distinction between causative faire and noncausative faire, and (c) in the production task the same person and number as in the interpretation task was not used. VanPatten (2002) also pointed to some problems in the Collentine (1998) study. To begin with, all subjects had received prior instruction on morphology. Secondly, Collentine had not provided explicit information to subjects and no incorrect processing strategies had been altered. Thirdly, there had been a big information load
in the activities. Finally, there had been no affective activities. On this basis, VanPatten concluded that “there are profound differences between the methodology of the replication studies and that of the original”, i.e. the VanPatten & Cadierno, 1993 study, (VanPatten, 2002, p. 785).

The essence of VanPatten’s reservations concerning the altering of learners’ processing strategies (which might be a limitation of IP) and the operationalization of PI and TI [i.e. they should be (a) pure PI or TI and (b) not simply input-based or output-based] seem in the right direction. However, although it is obvious that he had the right to criticize that the implementation of PI be according to his own and associates’ proposals (e.g. Lee & VanPatten, 1995), the same does not apply to TI.

In fact, the very term *traditional* can receive many interpretations. Grammar Translation and Audiolingualism are both traditional by now but instead VanPatten and Cadierno had chosen to limit the term to Paulston’s categorization (VanPatten & Cadierno, 1993, p. 230; VanPatten, 1996, p. 3 - 4) and on what “most second language textbooks published in the United States” proposed (VanPatten, 1996, p. 4). In all probability, Paulston’s classification is not the only viable alternative of traditional instruction available; consequently, it is open to a multiplicity of interpretations. Moreover, one cannot base a research paradigm on published textbooks especially with the plethora available. Just because they are available does not verify that they are valid and that they are all used. Things become more problematic when discovering that VanPatten and Cadierno had based the TI instructional packet on a single “best-selling” textbook, i.e. Knorre et al. (1985, qtd. in VanPatten & Cadierno, 1993, p. 230). At the same time, it is evident that although the rest of VanPatten’s critique deals with details, they do make a difference, especially the absence of affective activities in Collentine’s (1998) materials.
VanPatten and Wong (2004) attempted a replication study of Allen (2000) (VanPatten & Wong, 2004, p. 97). Their target structure was the causative form in French and the relevant IP Principle was the First-Noun Principle. After closely examining Allen’s (2000) materials supplied by Allen herself upon request, the 2 authors concluded that they felt “that the results that Allen received in her study are at least partly due to the differences in the content of the assessment tasks and in the treatment between the two studies” (VanPatten & Wong, 2004, p. 102).

The participants were undergraduates from 2 universities in the U.S.A. The participants from the 1st university attended a 4th-quarter course in French and the participants from the 2nd university attended a 3rd-semester course in French. They “would not have received” prior explicit instruction in the target form (p. 102). A 60% cut-off score was used. Participants were grouped in 3 groups: PI, TI, and C. Learning outcomes were assessed through interpretation and production tests. The results of the statistical analyses in the interpretation tests revealed significantly enhanced performance of the PI group of learners compared with the TI group of learners, whereas the C group showed no signs of improvement. The results in the production tests did not reveal any differences in the improvement between PI and TI, whereas the C group showed no signs of improvement. VanPatten and Wong discussed their results in cross-examination with Allen (2000) (VanPatten and Wong, 2004, p. 110-114). Finally, they emphasised the need for replication studies within SLA.

Farley’s (2004a) study was motivated by 2 interrelated issues. First, the replacement of the “mechanical practice” nature of TI as operationalized in previous IP studies “with a more meaning-based output” type of instruction (p. 143). Second, the investigation of a target form that was more intricate and hard than the target structures examined in previous studies, i.e. the Spanish subjunctive. The relevant
processing problems of the Spanish subjunctive, according to Farley, were the Lexical Preference Principle and the Sentence Location Principle. Collentine’s (1993) study on the Spanish subjunctive was a point of reference for Farley (Farley, 2004a, p. 144 - 146). Farley commented on the importance of the number of participants (p. 146) in order to reach definitive conclusions, a point which will be reiterated further along in the discussion part of the present thesis.

He also pointed out the importance of knowledge transfer “from exemplars in the treatment to novel test items” (p. 147). Despite the fact that knowledge transfer is a standard motivation behind any type of learning, it has somehow escaped research within SLA, Farley (2004a) being a notable exception. The fundamental fact that the knowledge learners acquire though instructional treatments ought to be transferable to different situations (the term is used instead of Farley’s novel test items since it is broader in scope) has been ignored. Apart from simply conducting research, there is no educational purpose behind not attempting to estimate whether what learners have acquired can be applied to different instances (through fine-tuned statistical analyses).

The final participant pool in Farley’s study consisted of 50 university-level students in a 4th-semester grammar review course in Spanish. The participants formed 2 groups: MOI and PI. These participants “did not receive explicit instruction or homework assignments on the Spanish subjunctive of doubt during previous weeks of the course” (p. 147 - 148). A 60% cut-off score on the pre-test and a number of cut-off measures were utilized. Learning outcomes were assessed through both interpretation and production tests. The analysis of the results exhibited some kind of enhanced learning outcomes in both the interpretation and the production data sets without any statistically significant differences (p. 155).
Farley’s discussion of the findings centered on the absence of mechanical activities in the MOI instructional treatment assuming that Collentine’s (1998) TI instructional treatment was not pure TI but rather “more MOI-like” (Farley, 2004a, p. 159). Farley’s (2004a) results closely resembled Collentine’s (1998) results (Farley, 2004a, p. 159). A second important observation is that Farley (2004a) discussed at length the nature of the target form as a causative variable for the procured results. However, citing the Cadierno (1995) and Benati (2001) findings, Farley (2004a) concluded that the causative variable was not the target form, but implied that the “operationalization of constructs” was the causative variable in his and Collentine’s (1998) study (Farley, 2004a, p. 162). Finally, Farley raised the issue of the beneficial effects of output and explicit information. As to the former, however, he abstained from theoretically justifying the role of output in SLA, correctly pointing out that SLA theorizing had not been able to consistently place a definitive role for output in second language acquisition (p. 163 - 164).

Collentine (2004) raised a number of issues concerning the statistical analyses used to procure results in PI research and other research within the FonF and FonFS research paradigm citing Norris and Ortega (qtd. in Collentine, 2004, p. 173 - 176). More specifically he pointed out that, “(i)f a carefully designed study does not control for the effects of an agent that could have sufficient importance on the outcome so as to compromise the goal of obtaining a sample that adequately represents the target population, the results are questionable regardless of whether the null hypothesis is accepted or rejected.” (Collentine, 2004, p. 176, parenthesis added).

There is a fundamental issue at stake here. More specifically, Collentine openly theorizes in favour of cut-off scores and / or the elimination of very low or very high outliers from data sets. He ignores the fact that the target population ought
to be carefully selected before conducting research, and not adapted to research purposes. It is in the latter case that the results are questionable, not in the former.

Collentine’s observation is motivated by the need to measure learning outcomes and procure homogeneous samples, overlooking the fact that deviations from the norm cannot be eliminated in a scientifically valid way. His observation emanates from the justified concern to ensure that external factors do not tamper with the results and that sample data sets are representative. However, the point is that it seems uncertain whether it is possible to control and validly assess all the possible external factors that might affect learners’ performance, such as individual differences, aptitude, memory, socio-economic background, etc. Even if such data are available to the researcher, it is impossible to calculate a host of other factors, such as weather conditions, psychological climate, time of day, and so on. The elimination of external factors in order to carry out accurate statistical analyses is the basis of Collentine’s statement; not the learners, neither learning per se.

Collentine’s observation is scientifically and statistically a valid one. However, it is not readily operationalizable, especially taking into account that the FonF / FonFS (and consequently PI) paradigms are relatively novel approaches, and despite the plethora of published studies in recent years, is in an infantile stage compared with more traditional approaches. In fact, as Vassilopoulos mentioned, “the assessment of a new biological parameter in a small sample does not guarantee a normally distributed sample, since the distribution of the parameter in the (whole) population is unknown” (Vassilopoulos, 1992, p. 54, parenthesis added, personal translation from the Greek original). PI is not a biological parameter but belongs to the field of “psychometric research”, according to Collentine (2004, p. 176) and can
be assumedly counted as a parameter in Vassilopoulos’s sense of the term *biological parameter*.

In plain terms, the research paradigms under investigation (such as FonF and FonFS) and the more state-of-the-art statistical analyses carried out (see e.g. Marsden and Chen, 2011) are based on the assumption of normal distribution which has not been thoroughly validated in the cases of such novel techniques. Also, an important point is overlooked, i.e. that “statistics deals with the probabilities about the validity of a hypothesis” and that “a hypothesis is never accepted or rejected with absolute certainty” as Vassilopoulos pointed out with emphasis (Vassilopoulos, 1992, p. 40, personal translation from the Greek original). Coupled with Collentine’s (2004) aforementioned observations, it needs to be emphasized that the learning outcomes estimated through statistical analyses are tentative assumptions. Despite recent debates concerning the appropriate statistical tools used, this fact merits more attention. Moreover, the statistical analyses of the results cannot be used as the one and only standard according to which learners’ performance is assessed.

To continue, based on VanPatten and Oikkenon’s (1996) findings that the causative variable for the learning outcomes of PI was not explicit information (EI) but the structured input (SI) activities, Wong (2004) aimed at examining “the generalizability of the findings of VanPatten and Oikkenon” (p. 188). The target structure selected was the change in the use of French articles in negative or affirmative statements and the principle examined was the *Lexical Preference Principle*.

The final participant pool consisted of 94 undergraduate students in a 1st-quarter course in French in a university in the USA. The participants’ L1 was English and they had not been exposed formally to the target structure. A 60% cut-off score
was used. The participants formed 4 groups, i.e. PI, EI, SI, and C. Learning outcomes were assessed through interpretation and production tests. As to the interpretation of the target structure, both the PI and the SI groups significantly outperformed the other 2 groups, without statistically significant differences between the PI and the SI groups. There were no statistically significant differences between the EI and C groups, either. As to the production of the target structure, the PI and the SI groups significantly outperformed the C group. Also, the PI group significantly outperformed the EI group. The SI and the EI groups did not exhibit significant in-between differences. This also applied to the EI and C groups.

Wong (2004) attributed the enhanced PI learning gains as to both interpretation and production to the SI activities (p. 196) and also emphatically summarized that

“(r)ecently, some SLA scholars remarked that among the focus on form techniques that are in the literature, PI appears to yield some of the most promising results (e.g. Carroll, 2001; Doughty. 2002; Ellis, 2002; Norris and Ortega, 2000) … According to these scholars, in L2 learning situations where input alone may not be sufficient, the best kind of intervention may be one in which input is structured so that learners can perceive and parse L2 stimuli more effectively to make accurate form-meaning connections (Doughty, 2002; Ellis, 2002).”

(Wong, 2004, p. 198, references in the original, parenthesis added)

Sanz (2004) investigated the effects of PI including 2 novel elements, i.e. (a) computer-implemented material and (b) the provision of explicit and implicit feedback. The target structure selected was “the Spanish O-cliticpro V S and O-cliticopro V sentences” (p. 246). The final participant pool consisted of 28 university students attending 1st- or 2nd-year Spanish courses in a university in the USA. They formed 2 groups: an “implicit feedback group” and an “explicit feedback group” (p.
A 60% cut-off score was used. There was no explanation phase before practice in either treatment condition. The explicit feedback treatment examined the “incorrect use of the strategy or VanPatten’s Principle 2” (p. 247), i.e. the *First Noun Principle*. The implicit feedback treatment attempted to “push learners to a different processing strategy” (p. 247). Learning outcomes were assessed through interpretation and production measures. As to interpretation, both groups exhibited significant gains. As to production, the structured input activities in isolation resulted in gains irrespective of feedback condition, i.e. implicit or explicit. Sanz concluded that “within PI it is positive evidence alone that leads to a change in processing and consequent L2 development” (p. 252).

VanPatten and Fernandez (2004) investigated the long-term effects of PI. The target structures selected were Spanish OVS sentences and clitic pronouns. The final participant pool consisted of 45 (p. 278) or 44 (p. 284) university students in a 3rd-semester Spanish FL course at a university in the USA. Only 1 PI group was formed. Learning outcomes were assessed through both interpretation and production tests, (a) immediately after instruction, and (b) 8 months after instruction. The statistical analyses as to both the interpretation and the production of the target structure revealed that despite decreased performance in the delayed post-test compared with performance in the immediate post-test, the learners’ delayed post-test performance was still higher to a significant extent than their performance in the pre-test.

VanPatten and Fernandez “believe the lack of a control group…does not constitute a serious problem” since such a control group did not show improvement in delayed posttests in similar studies”, such as Benati (2001), Cadierno (1995), VanPatten and Cadierno (1993), VanPatten and Wong (2004) (VanPatten & Fernandez, 2004, p. 285). However, since VanPatten and Fernandez (2004) did not
report the inclusion of a comparison control group, the results and the ensuing argumentation need to be interpreted with great caution.

Lee (2004) put forth a number of important observations pertaining to a certain extent to the majority of studies with the PI framework. Although he justifiably accepted the standard pre-test / post-test assessment measures and the ANOVA statistical analyses to assess learning outcomes, he made a number of comments, 2 of which are of interest at present. First, that “even a (cut-off) score of 60% indicates some prior knowledge, albeit incomplete prior knowledge” (p. 317, parenthesis added). Second, he argued that PI studies had focused on Spanish, French, and Italian with L1 English learners. He questioned, therefore, the generalizability of the findings to different L1 learner populations. In his own words:

“The PI research has examined three target languages, Spanish, French, and Italian, but only one native language, English. The homogeneity of the overall database lends weight to arguments of generalizability but is at the same time a limitation. Is Processing Instruction equally effective across a variety of native languages?”

(Lee, 2004, p. 318)

Furthermore, a prevalent point that Lee alluded to, without making explicit reference to it, was that the majority of learners used as subjects were university students who might have already had some kind of previous exposure to target forms (p. 317). This point, coupled with the fact that Spanish, French, Italian, and English are Romance languages (p. 321) led him to hypothesize that “(s)ome learners benefit more from PI than do others” (p. 318, 322). It should be reminded here that the 4 aforementioned languages share an almost similar alphabet and there is a large similarity between the L1 and L2 languages. This adds validity to Lee’s argumentation.
Marsden (2006) studied the acquisition of L2 French verb inflections for tense, number, and person through 2 classroom experiments. Her theoretical motivation was that at the time of the study “there ha(d) been no comparison of PI with an input-based approach that does not force learners to attend to the meaning of target forms” (p. 514, parenthesis added). Within the theoretical framework of the PI literature, Marsden focused on the *Lexical Preference Principle* and the *Availability of Resources Principle*.

The participants formed 2 groups: PI and EnI (enriched input). Since Marsden did not report the inclusion of a comparison control group, the results and the ensuing argumentation need to be interpreted with great caution. Marsden carried out 2 parallel experiments: the first experiment comprised of 27 13-14 year-old L2 French learners in a British secondary school in the U.K. They “had received approximately 180 hrs of classroom instruction in French and none had significant extracurricular exposure to French” (p. 525). These learners formed 2 groups: PI and EnI. The second experiment included 27 13-14 year-old L2 French learners in a second British secondary school in the same city of the U.K. 30 learners took part in the second experiment forming “Class 3 … (which) … carried out all the tests but did not have PI or EnI” (p. 538, parenthesis added). Marsden pointed out that “(t)he characteristics of the participants in Experiment 2 suggest that they might have had a slightly higher proficiency level than those in Experiment 1” (p. 540). Learning outcomes were assessed through many interpretation and production tests, in both oral and written form, comprising of various performance types, and focusing on all 4 productive and receptive skills: speaking, writing, reading, and listening. A pre-test, post-test, and delayed post-test research design was used.
The results of the statistical analyses of Experiment 1 showed that the PI Group exhibited significant gains between pre-tests and post-tests which were maintained in the delayed post-test. In Experiment 2 “the test-only class (something like a control group although not overtly stated as control by Marsden) made gains in all measures between the pretests and posttests despite the fact that there was no focus on the target forms between the pretests and posttests” (p. 541 - 542, parenthesis added). Marsden commented that

“(t)hese gains might have been due to a school-dependent test effect, due to the background in School 2 that emphasized sequences of grammar-focused activities and testing … To tease out any differences between the gains made by EnI and PI learners in Experiment 2, their learning gains (i.e. posttest scores minus pretest scores) were compared to those made by Class 3.”

(Marsden, 2006, p. 542, first parenthesis added, second parenthesis and emphasis in the original)

On this basis, Experiment 2 also failed to exhibit “learning gains in the EnI group” (p. 544), confirming, according to the author, the beneficial effects of SI in the PI literature.

Marsden discussed the details of her findings based on the numerous measures of performance she used, analyzing each measure separately. She reported 7 limitations, one of which is noteworthy: her study “used only Anglophone learners of a Romance language … (it) … cannot, therefore speak to the potential role of L1 in L2 input processing” (Marsden, 2006, p. 548, parenthesis added). This reiterates Lee’s previous argumentation that learners from different L1 backgrounds need to be incorporated in PI research (Lee, 2007, p. 318).

Morgan-Short and Bowden (2006) investigated input-based and output-based practice. Their target structure was Spanish preverbal direct object pronouns related to the First Noun Strategy. The theoretical framework of their study was the differential
effects of input and output in SLA. They chose PI as their input-based practice mode, and MOBI (Meaningful-Output-Based Instruction) as their output-based practice mode citing Swain’s *Output Hypothesis* as their focal point for MOBI (p. 38). They raised the necessity of “meaninffulness” as a crucial factor “in the effectiveness of instruction” following Cheng (2002) (Morgan-Short & Bowden, 2006, p. 37).

The final participant pool consisted of 45 1st-semester university students, aged 17-30 years old (mean age: 19.3 years). The L2 Spanish students formed 3 groups: PI, MOBI, and C. PI and MOBI were similar in the input they received but differed in whether the practice they received was input- or output-based. Morgan-Short and Bowden’s research agenda included 2 points worthy of mention:

1. “The control group was exposed to the target form and was approximately matched for time on task”, and
2. “an attempt was to control prior knowledge of the target form was made both by selecting first-semester Spanish students as participants and by using a pretest with a stricter criterion for inclusion (< 33%, which was chance level for the interpretation task.”

(Morgan-Short and Bowden, 2006, p. 44)

The fact that the comparison control group also received input on the target form is a serious limitation of this study, consequently, the results and the ensuing argumentation need to be interpreted with great caution. As to the issue of prior knowledge, the 33% cut-off score used was, probably, justified. However, it also operated as a measure of eliminating very low achievers from the data set, thereby implying that only relatively medium or high performers were included. If this line of reasoning were in the right direction, this means that the data set was modified so as to procure the desirable results. Therefore, it cannot be ascertained whether the treatment conditions could have procured beneficial or detrimental learning outcomes for all the participants irrespective of initial proficiency level as to the target form.
Learning outcomes were assessed through both interpretation and production tasks using a pre-test, immediate post-test, and delayed post-test (split-block) design. The results of the statistical analyses showed that while both the PI and MOBI groups outperformed the C group in the interpretation task, only the MOBI group outperformed the latter in the production task. The results did not exhibit significant differences between the 2 treatment conditions on both assessment measures.

Based on these findings, Morgan-Short and Bowden raised a number of issues. First, they explained the decreased MOBI performance in the delayed post-test as “a regression to the mean” (p. 52). Second, they discussed the improved MOBI Group performance on the basis of two lines of reasoning:

1. “input directly leads to acquisition but output might only affect acquisition indirectly or lead to skill specific knowledge”,
2. “both input and output can directly affect acquisition.”

(Morgan-Short & Bowden, 2006, p. 54)

As to the second line of reasoning, they maintained that “(i)n fact, the establishment and strengthening of form-meaning connections might be a common element that underlies the effectiveness of certain types of input-based and output-based practice” (p. 55 - 56, parenthesis added). Third, they attributed the C group subjects’ improved performance on (a) their exposure to the target structure, (b) test effects, (c) prior knowledge, or (d) a combination of the latter 2 points [i.e. points (b) and (c)] (p. 57). Fourth, Morgan-Short and Bowden explicitly referred to the issue of prior knowledge and cut-off scores (p. 57 - 58). More specifically, they pointed out

“the potential influence of prior knowledge of the target form in Spanish or of similar forms in other Romance languages due to participants’ prior language study … (which) … might have affected all treatment groups to some degree and would have been most visible in the gains of the control group.”
In the same vain, they attributed the difference in the results of their study and the results of similar studies to the fact that subjects in the similar studies “had more prior knowledge” (p. 58). The 2 authors also raised the issue of their subjects’ limited vocabulary knowledge as a limitation of their study. This limitation is valid and adds credence to the actuality that including very low proficiency L2 learners in similar research designs poses additional difficulties. Finally, Morgan-Short and Bowden argued that there is a role for both input-based and output-based instruction.

Toth (2006) cross-examined instructional treatments on the Spanish anticausative clitic ‘se’. The theoretical framework of his study was placed within the PI literature, the literature comparing PI and output-based types of instruction, and the literature related to the role of output in SLA. Toth’s choice of target structure, i.e. anticausative se, was made based on a number of characteristics that make it compatible with PI. First, it “derives a number of possible meanings in Spanish” (p. 332). Second, “the corresponding L1 (English) morphology varies considerably” (p. 332, parenthesis added). Third, due to its limited ‘communicative value’ (p. 333). Fourth, because its use “reflects Van Patten’s primacy of meaning principle and its corollaries” (p. 334), and more specifically the Lexical Preference Principle which Toth alluded to but did not mention explicitly (p. 334).

The participants formed 3 groups: PI, CO, (both including 55 native-adults speakers of English in total), and C (comprising of 25 speakers of English adults) (final subject pool: 80 2nd-semester students from 2 big, public universities in the U.S.A.). The PI and CO subjects were “enrolled in four intact sections of a … beginning L2 Spanish course” at one university and the “uninstructed” C group
subjects attended another university “where the L2 Spanish curriculum was almost identical to that of the first” (p. 331). A 50% cut-off score on the picture description part of the pre-test was used.

Toth’s research design can be summarized as follows: He compared 2 instructional treatments: (a) an “output-free PI” Group (p. 328) and (b) a “communicative output (CO) group” (p. 330, parenthesis in the original). Toth did not use the usual interpretation and production tasks to measure subjects’ performance, but instead learning outcomes were assessed through (a) production tests and (b) grammaticality judgement (GJ) tests. He replaced interpretation tasks with GJ tasks on the grounds that “superior PI outcomes on interpretation merely reflect the type of practice that learners received” following DeKeyser and Sokalski (1996) and Salaberry (1997) (Toth, 2006, p. 329). Assessment included quantitative analyses, and qualitative data “in the form of transcribed excerpts of classroom interactions to elucidate the issue of whether learner output in the CO group provides a similar linguistic environment to that of the PI group” (p. 330).

Toth summarised the results as presenting greater gains for the CO group than the PI group in production and similar results for both on GJ. His discussion of these results focused on the following points: First, he believed that the role of output is multifaceted and cannot be appropriately captured by VanPatten’s linear representation of the processes of SLA (p. 362), adding “that CO involved processing input and production” (p. 363), and also implying that “processing input and producing output are (not) two separate tasks” (p. 363, parenthesis added). He also postulated that “effective scaffolding techniques might be … (a) … key to any guidelines for beneficial output practice” (p. 363, parenthesis added). Secondly, he mentioned that the CO Group performance “on the production task suggests that
output may benefit acquisition in ways that go beyond” accidental structured input and enhanced input processing (p. 365). Third, he agreed with Izumi (2003) and Ellis (2003) that “the learner-internal language processes engaged during CO” can explain the greater gains of CO compared to PI (Izumi, 2003; Ellis, 2003, qtd. in Toth, 2006, p. 366). Fourth, he also explained the findings on the basis of Swain’s output hypothesis as to the meta-linguistic role of output (p. 368 - 371). All in all, he stated “that, just as Lee and VanPatten (2003) recommend, the best conditions for instructed learners are regular combinations of both structured input and meaningful output tasks” (Toth, 2006, p. 372, reference in the original).

Lee and Benati (2007) reported a number of studies study conducted by Lee and Benati. In the first study, the target structure selected was the Italian subjunctive of doubt and opinion. The relevant IP processing principles were the Selective Location Principle and the Lexical Preference Principle. Two groups were formed: PI and MOI. Instruction was both classroom-delivered and computer-delivered. According to Lee and Benati, “results showed no significant differences for mode of delivery on either instruction type. PI and MOI can be delivered effectively in both classrooms and computers” (p. 62). Lee and Benati also reported a second study conducted by Lee and Benati, where they “also tested the same variables using the French subjunctive of doubt”. They obtained the same results as for the Italian subjunctive. (p. 62).

Lee and Benati further reported a third (forthcoming at the time of publication) study (p. 167) conducted by Lee, Benati, Aguilor-Sanchez, and McNulty. The target structure selected was the Spanish preterite / imperfect aspectual distinction. The relevant IP processing problem was the Lexical Preference Principle. Three groups were formed. Instruction was both classroom-delivered and computer-delivered.
According to Lee and Benati (2007), “all three groups improved significantly after instruction and they all performed equally well” (p. 61). Lee and Benati also reported that the study of Lee, Benati, Aquilar-Sanchez, and McNulty “also tested the same treatment variables using Spanish negative informal commands” (p. 61). Lee and Benati reported that “Lee et al. obtained the same results as they did with aspectual distinction” (p. 62).

In Chapter 4, Lee and Benati presented the results of 2 studies (fourth and fifth in sequence) investigating the effects of PI on the acquisition of 2 grammatical forms: Italian adjective agreement and Italian subjunctive of doubt and opinion (p. 65 - 92). In the fourth study using as a target structure Italian adjective agreement, the processing problems examined were “the non-meaningfulness of gender agreement on adjectives, the redundancy of the adjective’s form, and possibly the location of the form in a sentence” (p. 67). The final participant pool consisted of 20 L1 English undergraduate students studying 1st-semester Italian at an L2 in a university in the UK. Two groups were formed: (a) structured input activities (SIA) and structured input activities plus input enhancement (SIAE). A 60% cut-off score was used. Learning outcomes were assessed through interpretation and production tests. As to both the interpretation and the production of the target structure, both groups performed equally well.

Lee and Benati concluded that

“(l)earners who perform structured input activities in which the target forms were enhanced aurally and visually did not make greater grammatical gains than learners who performed the same structured input activities but without the enhancements. Both groups of learners improved significantly as a result of instruction and they improved equally well.”

(Lee and Benati, 2007, p. 78, parenthesis added)
In other words, Lee and Benati considered SI as the causative variable of the enhanced learning outcomes and not input enhancement (IE). This can be considered as evidence that IE does not affect learning, or to be more accurate, does not summatively enhance the gains promoted by SI, i.e. it does not make any measurable contribution to activities which have been structured according to IP and PI guidelines.

Lee and Benati’s fifth study used as a target structure the Italian subjunctive of doubt. The relevant processing problems were captured by 3 PI principles, i.e. the Preference for Nonredundancy Principle, the Meaning-Before-Nonmeaning Principle, and the Sentence Location Principle. The final participant pool consisted of 24 L1 English undergraduate students studying a 1st-semester Italian course at the intermediate level. A 60% cut-off score was used. Three groups were formed: PI, PI plus enhancement (PIE), and PI plus enhancement delivered through computers PIE comp. Learning outcomes were assessed through interpretation and production tests. As to both the interpretation and the production of the target structure, the 3 treatment conditions procured similar gains.

Lee and Benati concluded that

“(i)t is the nature of the SI activities that makes the difference; enhancing the forms did not produce an additional effect … Enhancing the SI activities and delivering this treatment via computer terminal did not produce a greater impact on language learners’ performance”.

(Lee and Benati, 2007, p. 91, parenthesis added)

Gely (2005, qtd. in Lee and Benati, 2007, p. 97) had investigated the relative effects of PI and MOI, using as a target structure the French imperfect tense and focusing on the Lexical Preference Principle. The final participant pool consisted of
33 college-level students attending a 2nd-year French course. The participants formed 3 groups. PI, MOI, and C. Learning outcomes were assessed through interpretation and production tasks. Gely had reported that as to the interpretation of the target structure PI procured higher learning outcomes than MOI and C, and that, as to the production of the target structure, PI and MOI both procured equally higher learning outcomes (Gely, 2005, qtd. in Lee and Benati, 2007, p. 97).

Lee and Benati presented the results of a sixth study that investigated the relative effects of SI and TI using as a target structure the Japanese Past Tense, focusing on the *Lexical Preference Principle*. The participants were 27 beginner level students of L2 Japanese. Learning outcomes were assessed through interpretation and production tests. The statistical analysis showed that SI enhanced learning of the target structure more than TI.

Lee and Benati presented the results of a seventh “parallel” study that examined the relative effects of PI and MOI using as target structures (a) the Italian subjunctive of doubt and opinion and (b) the French subjunctive of doubt (p. 97 – 98). The participant pool for the Italian subjunctive consisted of 47 students divided in 4 groups: PI classroom, PI computers, MOI classroom, MOI computers. Learning outcomes were assessed through interpretation and production tests. According to the results, “the PI group performed better than the MOI groups in the interpretation task and both groups made equal gains in the production task” (p. 98). The participant pool for the Italian subjunctive consisted of 61 students. Lee and Benati reported that the results of the study on the French subjunctive procured the same results as the first study, i.e. the one investigating the Italian subjunctive of doubt and opinion.

Lee and Benati further reported an eighth study that investigated the relative effects of SIA (structured input only) and SIAE (structured input plus input
enhancement) using as a target structure the Italian Future Tense. The final participant pool consisted of 20 1st-semester English L1 students of Italian at the undergraduate level at a university in the UK. To be included, participants were required not to have had any prior exposure to or knowledge of the target form alongside a number of other criteria. In the SIAE practice, the verb endings were either pronounced more loudly (aural activities) or bolded and underlined (written activities). No time adverbial indicators were included in the activities in order to ensure that the attention of the learners was focused on “verb endings as the unique indicator of tense” (p. 103). Learning outcomes were measured through interpretation and production tests. The interpretation test was delivered in both oral and written mode and the production test was a written cloze passage requiring learners to supply the appropriate future tense form of the verbs in parentheses. As regards both the interpretation and the production of the target structure, both the SIA and SIAE “made equal gains” (p. 107, 108).

Lee and Benati concluded that

“(u)nenhanced structured input was equally as effective as enhanced structured input practice in helping learners to interpret and produce accurately the Italian future tense.”

(Lee and Benati, 2007, p. 109, parenthesis added)

Lee and Benati reported a study (ninth in sequence) by Lee and Benati that investigated the relative effects of SIA and SIAE using as a target structure the Japanese Past Tense marker. They “chose to target this form because it presents several processing problems of Italians learning Japanese as a second language” (p. 112). The final participant pool consisted of 26 adult, L1 Italian beginner learners of Japanese in a private school in Italy. A 60% cut-off score was used and participants
formed 3 groups: SIA, SIAE, and C. The practice did not include any time adverbials to ensure learners’ attention was focused on “verb endings as the only indicator of tense” (p. 115). In the SIAE practice, the verb endings were either pronounced more loudly (aural activities) or bolded and underlined (written activities). Learning outcomes were assessed through both interpretation and production tests. As to both the interpretation and the production of the target structure, similar results showed that “(t)he performance of the two structured input groups … (was) … superior to the control group’s but not different from each other” (p. 124, both parentheses added). Lee and Benati concluded that their “results underscore that SIA, no matter the way it is presented, is the main factor for learners’ positive performance” (p. 125).

Qin (2008) introduced a noteworthy novelty in studies conducted within the PI framework. More specifically, he aimed to assess PI (input-driven type of instruction) and dictogloss (DG) (an output-driven type of instruction). Qin’s study was theoretically placed in the field of studies comparing input-based and output-based instruction. The target structure was the English Passive Voice (limited to the Present Simple and the Past Simple Tenses due to the low proficiency level of the learners). The PI principle in question was the First Noun Principle. The Chinese participants were 7th grade secondary school children English L2 beginners whose age ranged between 13 and 15 years old and they belonged to 2 intact classes. Learning outcomes were assessed through comprehension and production tests. The pre-test and the post-tests (immediate and one-month delayed) procured dissimilar results. In the immediate post-test, the PI Group performed better than the DG Group as to both comprehension and production. Nevertheless, in the delayed post-test both groups produced similar results in both tests.
Qin argued that both PI (input-based) and DG (output-based) can efficiently aid the acquisition of the target forms by beginners. He insisted on a role for output in FLA and concluded that the metatalk produced by the learners in the dictogloss, as well as the meaningful nature of stories, aided learners’ comprehension as to the target form. Qin called for more research on PI and DG on a variety of target structures and was in favour of “a long-term grammar instruction curriculum” (p. 77).

VanPatten and Uludag (2011) investigated whether input-driven PI learning effects were transferable to output-oriented activities. The target structure selected was the English passive voice. The relevant processing strategy was the First Noun Principle. According to VanPatten and Uludag, “a lingering question in PI research … is to what extent the effects of PI are transferable to non-PI type tasks”, i.e. “on production tasks” (p. 46). The final participant pool consisted of 38 university Turkish L1 students of 2 classes in a public university in Turkey attending a 1st-year -English course. All the participants had taken a 1-year-EFL course and were adults (19 - 22 years old). Their proficiency level was pre-intermediate and they had not been exposed to the target form either through explicit instruction or homework prior to the implementation of the experiment. A 60% cut-off score was used. The participants formed 2 groups: PI and C. The PI instructional treatment included EI and 9 SI activities: 6 affective activities and 3 referential activities. The participants also received relevant feedback. Learning outcomes were assessed through interpretation tasks, sentence-level production tasks, and passage-reconstruction tasks. The statistical analyses revealed that as to all the task types, learners in the PI group made noteworthy gains, whereas learners in the Control did not exhibit any learning gains. VanPatten and Uludag pointed out the “(t)he effects of PI extended beyond the type of practice the participants received. The PI participants made significant gains on the
two measures of production included in the study” (p. 52, parenthesis added).

VanPatten and Uludag concluded that they “do not claim that input in general is sufficient, as free flowing input and input as used in other kinds of interventions (e.g. text enhancement, input flood, recasts) do not reveal robust and consistent effects we see with PI (Wong, 2005)” (VanPatten & Uludag, 2011, p. 52, both parentheses and reference in the original).

3.5.3. Processing instruction: research: summary

It seems that studies on processing instruction can be divided in 3 categories. The first category includes studies comparing the learning outcomes of PI-instructed groups compared with TI-instructed groups. The most significant finding of such studies was that whereas TI groups procured enhanced learning outcomes only on production measures, PI groups procured enhanced learning outcomes on both production and interpretation measures. The second category includes studies comparing PI-instructed groups with output-instructed groups. The findings of these studies were different. Some studies reported beneficial effects similar to the ones of the first category, whereas others reported that PI groups procured enhanced learning outcomes only on interpretation measures unlike output-instructed groups which either procured enhanced learning outcomes only on production measures, or on both production and interpretation measures. In fact, unlike the studies in the first category, there is no common pattern as to the findings of these studies. This is partly due to the fact that output-based instruction was not implemented in the same way in most of these studies, unlike the studies including TI groups which were more uniformly operationalized. The third category includes more recent studies in which a range of
different techniques were incorporated such as input enhancement and dictogloss. Also, this category includes studies with diverse aims such as knowledge transfer. In fact, these more recent studies are so different both as to their research design and their aims, that it impossible to discover a common pattern.

3.6. Processing Instruction & the English Past Simple Tense

3.6.1. Processing Instruction & the English Past Simple Tense: introduction

PI-centered research using as a target structure the English Past Simple Tense includes a variable almost absent from the rest of the studies within this research paradigm: young learners as participants. Another difference in relation to other PI-related studies concerns the L1s of these young learners. More specifically, most PI studies have included adult, university-level English L2 learners whose L1 was usually Spanish or French. In contrast, Benati’s (2005) participants are young Greek L1 and Chinese L1 EFL learners. The present section includes only this study.

3.6.2. Processing Instruction & the English Past Simple Tense: research

Benati (2005) conducted a study involving Greek L1 and Chinese L1 learners of English. His aim was to investigate the English Past Simple Tense because its acquisition was problematic for both groups of learners (p. 76). Benati’s study can be placed within the wider theoretical framework studying input-based instruction (Benati’s PI group) and output-based instruction (Benati’s TI and MOI groups) (p. 68 - 73). The PI principle in focus was the Lexical Preference Principle.
The participants were (a) “47 Chinese students (original pool 52 subjects) studying English in a secondary school in China (all at the age of 12-13)”, and (b) 30 subjects “in total”, “studying English in a secondary Greek school (all at the age of 12-13)” (p. 74 - 75, parentheses in the original). They were grouped randomly under 3 different treatments: PI, TI, and MOI (meaning-based output instruction). The participants were English L2 learners without any prior knowledge of the target form. Since Benati did not report the inclusion of a comparison control group, the results and the ensuing argumentation need to be interpreted with great caution.

Learning outcomes were assessed through interpretation and production tests. A pre-test and an immediate post-test were used to measure performance. Both Greek and Chinese learners in the PI group performed better than the TI and MOI Groups in interpretation, whereas in production tasks all 3 groups had the same performance. Benati’s findings then closely replicated VanPatten and associates’ findings and Benati’s study was yet another instance where input-based instruction proved to be better than output-based instruction (p. 83).

Benati explained the results on the basis of (a) the nature of the target structure and (b) the difficulties it poses to learners of different L1 backgrounds, especially the Chinese ones. In his conclusion, he focused on 4 points:

1. the beneficial effects of PI as to the alteration of the processing strategies of the learners.
2. output-based practice should succeed input-based practice.
3. SI is the reason behind improved PI outcomes.
4. “PI is an effective instructional treatment across different native languages (Chinese and Greek) and a different population (school learners).” (Benati, 2005, p. 84 - 85, parentheses in the original)
He concluded by stating the need for production measures that are time-pressured.

3.6.3. Processing Instruction & the English Past Simple Tense: summary

It is not valid to assess the effects of processing instruction as to the acquisition of the English Past Simple Tense on the basis of a single study, i.e Benati (2005). However, a number of observations related with this study can be made. The first concerns the L1s of the participants, i.e. Greek and Chinese. This is in sharp contrast to the majority of studies on processing instruction which have Spanish and French L1 participants. The second observation concerns the age of the participants. More specifically 11- to 12-year-old participants are almost absent in research on processing instruction which has mainly focused on adult university-level participants. All these factors combined render any definitive conclusions impossible.

3.7. Structured Input: Research

3.7.1. Structured input: research: introduction

Processing instruction includes the presentation of explicit information on erroneous processing strategies as well as practice in the form of structured input activities. A number of studies have attempted to investigate whether these activities are the causative variable of enhanced learning outcomes reported by processing instruction. In plain terms, the presentation phase of PI has been omitted. This deconstruction of the constituents of PI has led to the more implicit-like type of
instruction of SI. Structured input practice comprises of referential and affective activities.

As a matter of fact, more recent research has further de-contextualized the original PI teaching methodology. For example, Marsden and Chen (2011) presented in another section of this chapter, have attempted to cross-examine the learning effects of these 2 types of activities. Furthermore, research on SI has incorporated visual and textual enhancement formats (see e.g. Lee & Benati, 2007, reviewed in the previous section).

However, SI-centered research has downplayed an important methodological factor, i.e. that SI activities include input flooding on the target structures investigated. In this respect, enhanced learning outcomes reported by exposure to SI activities may not be an attribute of the structuring of activities per se, but a logical consequence of exposure to a high number of target forms embedded in such activities. In order to validly assess whether SI activities, and not the embedded input flooding in such activities, are responsible for the enhanced learning outcomes, the following research design seems appropriate: The cross-examination of 3 instructional treatment groups: (a) a standard SI group, (b) an input flood group exposed to exactly the same tokens as the SI group in the form of reading passages, and (c) a control group without any kind of exposure to the target form. The main study of the present thesis reported in chapter 4 incorporates such a research design.

3.7.2. Structured input: research: studies

VanPatten and Oikkenon (1996) were the first to investigate structured input (SI) as the causative variable of the PI-centered research agenda. They investigated
whether the explanation phase or the SI activities were responsible for the results of previous studies. Three groups were formed: (a) PI as in VanPatten and Cadierno (1993), (b) explanation only, no structured input activities (Eonly), and (c) SI activities only, no explanation (SIonly). Learning outcomes were assessed through the same assessment tests as in VanPatten and Cadierno (1993) with only 1 post-test 1 day after instruction. As to the interpretation of the target structure, both the PI group and the SIonly group showed significant gains, unlike the Eonly group that remained stable. As to the production of the target structure, all 3 groups improved but only the PI group and the SIonly group showed statistically significant improvement from pre-test to post-test. VanPatten and Oikkenon (1996) concluded that SI was the causative variable responsible for the learning outcomes in that and previous similar studies.

Erlam (2003) investigated the acquisition of direct object pronouns in L2 French (with the exception of ‘en’ and reflexives). Her theoretical paradigm belonged to the literature linked with input-based instruction. According to Erlam, input-oriented instruction can be either implicit-oriented or explicit-oriented. The former includes (a) “enriched-input instructional techniques” [i.e. input flooding (Trahey, 1996, Trahey & White, 1993, qtd. in Erlam, 2003, p. 560)] and (b) “enhanced-input instructional techniques” [i.e. typographical enhancement (Alanen, 1995; Doughty, 1991; Leeman, Arteagoitia, Fridman, & Doughty, 1991 qtd. in Erlam, 2003, p. 560)]. The latter includes input-processing instruction and structured-input instruction (p. 560).

Erlam used the following research design: Her instructional treatments were a structured input (SI) group and an output-based instruction (OBI) group, as well as a C Group. SI was based on VanPatten and Cadierno (1993b), while the OBI group was modelled loosely on what Dekeyser (1998) had defined as cognitive code including
explicit instruction and form-focused activities (Erlam, 2003, p. 568). The C group was not exposed to the target structure at all, apart from the assessment tests (p. 569). The final participant pool consisted of 70 participants (around 14 years old) belonging to 3 classes of 4th-form students in a secondary school in New Zealand. English was the native language for the majority of subjects. The participants had not received any prior instruction on the target structure. Also, they were introduced to it “earlier than scheduled in the syllabus plan” (p. 566). Learning outcomes were assessed through both comprehension and production measures in all 4 skills: listening comprehension, reading comprehension, written production, and oral production.

Despite the complicated findings due to the variety of skills assessed, Erlam reported a significant overall tendency for the OBI group to outperform the SI group. Both groups made greater gains than the C group in both comprehension and production tests. Erlam attributed the increased gains of the OBI group to 3 factors:

1. the meaningful nature of the majority of practice activities,

2. the attentional focus on the target form for both groups, enhanced by the inclusion of feedback, and

3. the fact that output-oriented instruction assisted automatized utilization of the target form.

(Adapted from Erlam, 2003, p. 577 - 578)

Finally, Erlam pointed to the necessity of including “language measures (that) require a pressured response” (p. 579, parenthesis added).

Despite the fact that response-pressured assessment measures have been identified as indicators of implicit language learning, it remains unanswered whether
they can be implemented with young and/or beginner learners. Without doubt, they are a useful tool when employed with older and/or more proficient learners and if appropriately designed, are probably the best tools of tapping on (implicit) language learning. However, there is an unresolved problem lingering. It concerns the fact that different research designs have used thus far a large number of assessment tools and this is definitely sound both theoretically and methodologically, since researchers utilize tasks according to the aims of their studies. The problem is related with the reality that this multitude of measures renders the different studies very difficult to compare. In fact, the use of different assessment measures results in non-comparable findings. It is high time SLA research found a way out of this quandary. It is one of the outstanding merits of research studies conducted within the IP and PI paradigm that the interpretation and production tests have remained almost unchanged up to the present day. In this way, the authors of these studies are validated in making comparable assumptions and reaching tentative overall conclusions based on a number of similar studies. On the other hand, research studies outside IP and PI have rarely attempted to utilize the same methodology to measure performance. However, the authors of these studies insist on cross-examining their findings overlooking the fact that they have measured different things.

Benati (2004) also examined the SI phase of PI. The target structure selected was the Italian future tense and the principle examined was the *Lexical Preference Principle*. The final participant pool consisted of 38 subjects, 2nd-semester undergraduate students of Italian at a university in the UK. Three groups were formed: PI, SI, and EI. The assessment measures included interpretation and production tests. The statistical analyses revealed that as to both the interpretation and the production of the target structure, the PI and SI groups both improved to a similar
extent, whereas the EI group exhibited a small improvement. Benati concluded by assuming that whereas explicit information might be necessary for other types of instructional treatments, it is not necessary for structured input treatments (p. 217). Benati did not report the inclusion of a comparison control group so the results and the ensuing argumentation need to be interpreted with great caution.

Farley (2004b) investigated the relative effects of the SI phase within the PI paradigm, partially replicating VanPatten and Oikkenon (1996), (Farley 2004b, p. 228). The target structure selected was the Spanish subjunctive and the PI principles explored were the **Lexical Preference Principle** and the **Sentence Location Principle**, just like Farley (2004a). The final participant pool included 54 university students in a 4th-semester Spanish course. They formed 2 groups: PI and SI. A 60% pre-test cut-off score was used “to avoid ceiling effects” (Farley, 2004b, p. 228). Learning outcomes were assessed through interpretation and production tests. As to both the interpretation and the production of the target structure, the PI group performed better than the SI group.

Farley (2004b) attributed these results as due to the nature of the target form. He believed that “unlike tense… and agency…, mood as a feature is not readily understandable or clear to learners” (p. 237). “However, because the SI group did make significant gains without explicit information, we cannot conclude that explicit information is necessary” (p. 238). Farley’s observation concerning the nature of the target form as a causative variable for the findings does make sense. However, his argument about the non-necessity of explicit information and his ensuing discussion are counter to the findings of his study.

As aforementioned, Lee and Benati (2007) after conducting a series of studies including a variety of target forms concluded that their results constituted evidence
that it does not matter whether SI activities are presented using input enhancement (IE) formats or not, but that the SI activities were the causative variable of the findings in their numerous studies (Lee & Benati, 2007, p. 125).

In a laboratory setting using computers, Fernandez (2008) investigated the relative effects of PI [explicit information (EI) + structured input (SI)] and SI [structured input only (SI only)] in 2 experiments using a different target structure each time. Her aim was “to observe, through online treatments, whether explicit information assists acquisition in a way that ha(d) not been measured in previous processing instruction (PI) studies” (p. 277, first parenthesis added). Her research agenda was placed within the framework of the PI literature examining whether the positive effects of PI are to be attributed to EI or SI. In the first experiment, Spanish OVS-type sentences with clitic object pronouns in the context of the First-Noun Principle were examined. In the second experiment the target structure was the Spanish subjunctive in expressions of doubt examining the Lexical Preference Principle and the Sentence Location Principle.

The subjects were 3rd-semester college L2 Spanish learners at the beginner or intermediate level. The measures applied to assess learning were 3: (a) trials to criterion, (b) response time, and (c) accuracy after criterion. In the first experiment, no difference was found between the PI and SI Groups. However, in the case of the subjunctive as a target form, the results were different: the PI Group managed to process the target forms “sooner” and “faster” (p. 277).

Fernandez offered a number of explanations for the results of her experiment. She noted that the findings indicated that EI appeared to be beneficial for the correct processing of the subjunctive, and believed that “the benefits of explicit information might depend on the nature of the tasks (on-line tasks) and the processing problem.”
itself (p. 277, parenthesis added, see also p. 297). Fernandez claimed that “EI seems to be beneficial when the task is to notice and process a single form, but it seems not to play any role when the task is to assign different grammatical roles in sentences” (p. 298). According to Fernandez, explicit information (a) could help the noticing of otherwise unnoticed forms, “or that might have taken more trials to notice”, but it did not “obviate or reduce the need for minimal exposure time”, and (b) it did have an effect “in building a new processing strategy” (p. 298).

Henshaw (2012) investigated the relative effects of referential and affective SI activities. The target structure selected was the Spanish subjunctive with adverbial clauses of time. Both instruction and practice were computer-based. The relevant processing strategies examined were the Preference for Nonredundancy Principle and the Sentence Location Principle. According to Henshaw, referential activities exhibit task-essentialness (term used by Loschky & Bley-Vroman, 1993, and Tomasello & Herron, 1989, qtd. in Henshaw, 2012, p. 396), whereas “(i)n contrast, affective activities lack task-essentialness” (p. 396, parenthesis added). However, the author immediately affirmed that “the notion of combining both referential and affective activities may still be in line with both Schmidt’s and Bley-Vroman’s claims” (Henshaw, 2012, p. 397, references in the original).

The participant pool consisted of 103 university English L1 students in a university in the USA, attending 2 4th-semester Spanish courses. “During the data collection period, the use of the subjunctive with adverbial clauses was not formally presented in either of the courses” (p. 401). The participants were

“late learners of Spanish as a foreign language (average age of first exposure was 13) … (who) … had less than one semester of immersion in a Spanish-speaking country, and less than 2 semesters of other foreign language study. Thus, ensuring no knowledge of the subjunctive in other languages.”
A 33% cut-off score was used. The participants formed 3 groups: RA (referential + affective activities), REF (referential activities), and AFF (affective activities). Since Henshaw did not report the inclusion of a comparison control group, the results and the ensuing argumentation need to be interpreted with great caution. Practice was coupled with the provision of feedback. Learning outcomes were assessed through interpretation and recognition tests. The format of the recognition tests “mimics that of controlled sentence-level production, except that learners are provided with options rather than being asked to produce the form” (p. 401).

The statistical analyses procured a number of results. First, the standard deviation values were very high in the post-tests. Second, as to the recognition of the target form, all 3 instructional treatment groups “made equally significant gains from pretest to immediate posttest” (p. 407) and “delayed posttest scores were still significantly higher than pretest scores for all treatment groups” (p. 408). Third, as to the interpretation of the target form, the results resembled the results of the recognition tests with the additional fact that “in the case of interpretation items, results revealed that the two groups that engaged in affective SI activities (RA and AFF) were able to better maintain learning gains over the span of two weeks than the group that completed only referential SI activities (REF)” (p. 409-410, both parentheses in the original).

Henshaw reported that “all participants received nonparadigmatic information in English about how the subjunctive is formed, where it is located within a sentence, and when it is used in adverbial clauses of time … (and) … were warned about a
particular processing strategy that might negatively affect their acquisition of the target form…” (p. 402 - 403, parenthesis added).

The author provided a number of explanations for the results. First, that “the same amount and type of “EI offered to all groups is the causative variable for their learning gains” (p.410). Second, EI “followed by SI practice, regardless of its task-essentialness”, might have been another causative variable for the learning gains (p. 410). Third, “(t)he positive evidence provided by affective SI activities appears to have been as beneficial to learners as the negative evidence that the corrective feedback of referential activities may supply” (p. 410, parenthesis added). Henshaw commented that there is a need to examine whether SI activities can affect “learners ability to generalize to new instances” (p. 410).

3.7.3. Structured input: research: summary

Two are the most important issues that SI-related research needs to address. The first is the issue of control conditions. With the notable exception of Erlam (2003), none of the studies reviewed in this section have included a control group. Consequently, the enhanced learning outcomes of SI groups are to be interpreted with great caution. The second issue concerns the amount of input offered to learners through SI activities. In other words, since SI groups are no longer presented with the explicit information on processing strategies standard PI groups receive, research on SI needs to promptly address this issue. This issue has received relatively little attention in the pertinent literature.

3.8. Structured Input & the English Past Simple Tense

SI-centered research using as a target structure the English Past Simple Tense includes a variable almost absent from the rest of the studies within this research paradigm: young learners as participants. Another difference in relation to other PI-related studies concerns the L1s of these young learners. More specifically, most PI studies have included adult, university-level English L2 learners whose L1 was usually Spanish or French. It is surprising that Marsden and Chen’s (2011) participants are Taiwanese L1 EFL learners. The present section includes only this study.

3.8.2. Structured input & the English Past Simple Tense: research

Marsden and Chen (2011) investigated the relative effects of referential and affective activities included in PI materials. The target structure they examined was the Past Simple Tense -ed verb inflection. Within this framework, they tried to specify whether implicit or explicit knowledge were sustained by referential or affective tasks. They argued that affective and referential activities are quite different, arguing that “affective activities provide positive evidence about the target feature, whereas referential activities provide both negative and positive evidence” (p. 1061). Discussing the way affective activities might impact learning, Marsden and Chen proposed that affective activities might operate as an input flood in some cases. Despite accusing VanPatten as “agnostic about the precise knowledge generated by PI and whether awareness is involved in learning PI”, Marsden and Chen commented
“that the observed learning gains following PI have often been interpreted as evidence of knowledge that tends to be implicit rather than explicit” (p. 1063). Consequently, one of their primary aims was to examine what kind of learning is promoted by referential and affective tasks.

The participants were 120 Taiwanese EFL learners in the 6th-grade in a Primary School in Taiwan. Participants were about 12 years old. The regular Past Simple Tense -ed inflection was chosen as a target structure because (a) the participants had minimal or no exposure to it, and (b) the acquisition of this target form had been reported as presenting problems to English L2 learners by Benati (2005) (Marsden & Chen, 2011, p. 1066). The participants formed 4 groups: (a) Referential + Affective, (b) Referential only, (c) Affective only, and (d) C. Learning outcomes were assessed through numerous tests. The results of the statistical analyses conducted on the various tests showed that even though learning was enhanced by the referential activities, participants did not show improvement after exposure to the affective activities. Furthermore, the analyses implied that referential activities might exhibit traits of explicit knowledge learning.

As to the issue of whether explicit or implicit knowledge is tapped through referential or affective activities, Marsden and Chen admitted that tests of implicit knowledge were not exhaustive and, in fact, they abstained from claiming that implicit knowledge might have been promoted. In their own words: “We acknowledge that our data do not warrant commentary on whether initial learning that is shaped by explicit knowledge can later become part of a system that is inaccessible to consciousness” (p. 1088).

In sum, Marsden and Chen did not find any enhanced implicit knowledge learning and their findings also failed to promote enhanced learning outcomes of the
target structure by the affective activities. However, they seem to have undervalued the fact that implicit knowledge cannot usually be explicitly verbalised contrary to explicit knowledge (Hulstijn, 2002, p. 205; Ellis, R., 2006, p. 437), and they might have disregarded the fact that explicit and implicit knowledge are, according to some researchers, a continuum, (Dienes & Perner, 1999, qtd. in Ellis, 2004, p. 232; Bialystok, 1990; Schmidt, 1994a, qtd. in Erlam, 2006, p. 464; see also Doughty & Williams, 1998, p. 5 for a similar FonF and focus on formS continuum) and not polar opposites.

Marsden and Chen cited Benati and Lee (2008) who had reported that the learning outcomes of structured input activities (SIA) may extend to learning of forms beyond the target structures the SI treatment focuses on (Benati & Lee, 2008, qtd. in Marsden and Chen, 2011, p. 1086). They also noted that Marsden (2004) included in her practice activities that could be considered as combining both referential and affective activity type elements (Marsden, 2004, qtd. in Marsden and Chen, 2011, p. 1092).

3.8.3. Structured input & the English Past Simple Tense: summary

Despite the aforementioned reservations as to the Marsden and Chen (2011) study, it is a fact that SI, as operationalized in their study, did have an effect as to the learning of the target structure. Moreover, the 2 authors alluded to the fact that affective activities might operate as input flood, an issue raised in the introduction to this section. Nonetheless, in their detailed analysis of the attributes of both referential and affective activities they did not mention that both types of activities include a large number of target structure tokens, i.e. both include input flooding. Finally,
Marden and Chen discussed the inferior learning outcomes of affective activities downplaying their learner-friendliness and focusing only on their measurable learning outcomes. However, affective activities are the only component of PI and SI that is the most learner-friendly.

Conclusion

This chapter presented both the theoretical frameworks and the research implementations of textual and visual input enhancement, input flooding, input processing, processing instruction, and structured input. An attempt was made to discover common patterns in the research reviewed, but this was not possible due to variations in the research designs. It seems that only older studies on processing instruction comparing PI and TI formed a category susceptible to the discovery of a common pattern in the data. More recently, researchers in the aforementioned branches of SLA have started investigating a host of factors making the discovery of common patterns as to the findings even more difficult.

As a concluding remark, it needs to be mentioned that very few of the studies reviewed in this chapter have included young learners as participants. The participants of the main study of the present thesis lacked the necessary linguistic capacities to benefit from explicit rule instruction. Consequently, it was deemed necessary that they were exposed to input enhancement techniques only, namely textual input enhancement in the form of bolding in combination with an input flood on the English Past Simple Tense –ed in the IF group and, for the same reason, the most implicit-like element of processing instruction, i.e. structured input, was incorporated in the research design. It was hoped that this research design would aid and shed light on the
acquisition of the regular English Past Simple Tense –ed. It was also hoped that such a research design would answer the research questions and confirm or not the research hypotheses.
CHAPTER 4:

METHOD

Introduction

This chapter begins with a presentation of the research questions and hypotheses of the present thesis. In this respect, it reports on the implementation of 2 full-scale studies, i.e. a pilot study and the main study, that were carried out to examine the effects on acquisition of implicit instructional treatment methodologies and techniques, i.e. SI and IF & TIE, by young Greek L1 EFL Primary State School learners. The target structure of both studies was the regular English Past Simple Tense –ed. First, the teaching methodologies and techniques employed in both studies are briefly explained. Second, the research design of both studies is presented. Third, the profile and role of the instructor / researcher in both studies is described. Fourth, the profile of the school where both studies were carried out is outlined. Fifth, the participants of both studies are presented. Sixth, the research design of both studies is reported. In the seventh section, the pilot study is described. This section includes the following sub-sections: (a) short introduction, (b) tokens, (c) materials and implementation, (d) research schedule, (e) scoring, and (f) short summary. The eighth section illustrates in detail the changes that were incorporated in the main study based on the experience drawn from the pilot study. The ninth section describes the main study and includes the same sub-sections as the seventh section, i.e. the pilot study.
The chapter concludes with some final remarks pertaining to the relationship between the main study experiment and the research question and hypotheses.

All the supplementary materials are available on 2 separate CDs. The 1st CD (CD 1) includes all the pilot study and main study materials, along with the parental consent forms required in both studies, except for the photographic materials. The 2nd CD (CD 2) includes all the main study photographic materials.

4.1. Research Questions and Hypotheses

The present thesis aimed at answering 2 main research questions:

1. Are the SI and IF & TIE implicit methodologies appropriate for the acquisition of the regular English Past Simple Tense –ed target structure?

2. Are the SI and IF & TIE implicit methodologies appropriate for the acquisition of the target structure by young Greek L1 EFL learners?

On this basis, the 2 hypotheses underlying the main study were formulated as follows:

1. Implicit instruction (in the form of the SI and IF & TIE instructional methodologies) can help young Greek L1 EFL learners acquire the regular English Past Simple Tense –ed structure.
2. Young Greek L1 EFL learners learning the regular English Past Simple Tense -ed after being subjected to the 2 different implicit instructional treatments will benefit the most from the instructional treatment that is the most explicit-like and both will outperform a 3rd comparison, control group. More specifically, young Greek L1 EFL learners learning the regular English Past Simple Tense -ed after being subjected to the more explicit-like implicit instructional treatment of structured input (SI Group) will benefit more than young Greek L1 EFL learners learning the regular English Past Simple Tense -ed after being subjected to a more implicit instructional treatment combining input flooding and textual input enhancement (IF Group). Both groups will outperform the 3rd group (C group, control) that will take part only in the assessment measures without any kind of exposure to the target structure.

4.2. Research Methodology

Both the pilot study and the main study instructional treatments were based on 2 proposals about implicit types of teaching / learning: structured input (SI) and input flood (IF). The former is based on the premise that second or foreign language input needs to be structured in a way that facilitates its subsequent processing by the learners in order to enhance language development and acquisition. The latter is based on the premise that maximizing the amount of exposure to foreign language input enhances language development and acquisition. Both proposals are input-based and, in the case of EFL settings like the present one, presuppose that, in the limited classroom time available, target structures are pre-selected and relevant teaching and learning materials are carefully designed. Although SI sides with foreign language
teaching/learning contexts aiming at structuring the input in a way that inherent rules and patterns are processed and subsequently acquired and IF draws on naturalistic language teaching and learning settings in which abundance of input leads to unconscious and implicit acquisition, both are implicit interventions without any kind of rule presentations or explanations provided. It is difficult, therefore, to accurately specify the amount of input required to foster acquisition in each case since the contexts/Settings differ in the way input is provided. Nonetheless, due to reasons of research requirements, the amount of input offered to participants in both instructional treatments was accurately balanced in both the pilot study and the main study.

Textual Input Enhancement (TIE) was also incorporated in the design of both the pilot study and the main study IF materials. More specifically, in the pilot study IF instructional treatment materials the target structure tokens were underlined, whereas in the main study IF instructional treatment materials the target structure tokens were bolded. In other words, a different TIE technique was used in each case. The aim behind using TIE techniques in the IF instructional treatment materials was to increase the perceptual salience of the target structure tokens and enhance subsequent noticing based on the premise that enhanced noticing would promote acquisition of the target structure on the part of the young learners. Since the participants were young learners, the inclusion of TIE was incorporated in order to induce them to pay attention to the target structure and avoid unnecessary distraction on linguistic features other than the target structure. The decision to use a different TIE technique in the main study was also motivated by the decision to assess the learning effects of yet another TIE technique.

Another pertinent difference between the pilot study and the main study was that in the pilot study the target structure tokens were underlined in 2 SI referential
activities, as well (2\textsuperscript{nd} and 4\textsuperscript{th} pilot study SI activities, see subsection 4.7.3.1). In other words, in the pilot study both the SI and IF instructional treatments incorporated TIE in the materials. In the main study SI instructional treatment materials the target structure tokens were not textually enhanced in any way. In other words, TIE was not incorporated in the main study SI instructional treatment.

The motivation behind this decision was based on the assumption that, since SI as a teaching technique lies on the foundation that input has already been structured so as to promote subsequent processing and acquisition, textually enhancing the target forms might counteract with this input structuring, operating as a distracting factor. In other words, SI coupled with TIE in the pilot study might have negatively tampered with internal input processing on the part of the young learners by forcing them to focus their attention on the form of the target structure rather than on meaning.

All in all, the research methodology of both the pilot study and the main study were based on the teaching methods of SI and IF, coupled, in one way or another, with TIE techniques. Despite the fact that the changes in the main study based on the pilot study are reported in detail in section 3.8., it was deemed necessary to include in the present section the differences in the TIE techniques between the 2 studies in order to introduce and describe more accurately the methodological foundations of the whole research project.

4.3. Instructor

The instructor and researcher of the both the pilot study and the main study, and author of the present thesis, were the same person. It was hoped that this would aid in avoiding any possible effects caused by different teaching styles or techniques. The
alternative procedure, i.e. using different instructors, would require teacher training sessions; this was not possible due to time constraints. A possible limitation, of course, is the so-called researcher expectancy, i.e. influencing learner performance along the lines of the researcher’s predictions for the experiment. Even though great attention was paid in trying not to influence the learners on the basis of my expectations, I cannot claim that researcher expectancy had absolutely no effect on the results since no objective evaluation of such an effect was made. This ought to be taken into account when interpreting the results.

4.4. School Profile

Both the pilot study and the main study were implemented at the 3rd Model Experimental State Primary School of Evosmos which belongs to the Directorate of Primary Education of Western Thessaloniki, Prefecture of Thessaloniki, Greece. This Primary School is affiliated to the Department of English Language and Literature of the Aristotle University of Thessaloniki, Thessaloniki, Greece. This affiliation presupposes and ensures that English as a school subject forms a more focal part of the curriculum than in mainstream Primary Schools. Also, many research projects, especially concerning English, are run throughout the year mostly organized in collaboration with the university. English is taught for more hours than in mainstream Primary Schools and is also used for cross-curricular content-based approaches, e.g. some subjects are taught in English. There are more English teachers available and this also enables smaller numbers of pupils in English classes. These pupils are taught with a range of methodologies and approaches. To enter the 3rd Model Experimental State Primary School of Evosmos, enrolment applications are compulsory and, due to
the large number of applications, selection is based on drawing lots. Consequently, pupils attending this school are not representative of the total Primary School pupil population throughout the country of Greece. Their majority belongs to the wider pupil population of metropolitan Thessaloniki, Greece. English teachers are likewise not typical of mainstream English teachers in Primary Schools. They are State-appointed and selected after applying to a separate authority responsible for (Model) Experimental State Primary Schools, i.e. ΔΕΠΠΣ, on the basis of predefined criteria for a 5-year term of office, during which they are regularly assessed.

4.5. Participants

The participants of the pilot study were 3rd grade pupils who were grouped in 3 separate groups (SI, IF, and C), aged 8 years old. Parental consent requirements (pilot study parental consent form available in CD 1, Appendix 1) limited the total number of participants to 32: SI: 11, IF: 11, and C: 10. The pilot study participants were not randomly assigned as they were grouped mostly according to their regular English class groupings which meant that they had not been uniformly instructed in English.

The participants of the main study were also 8-year-old 3rd grade pupils, who were grouped in 3 separate groups (SI, IF, and C). Parental consent requirements (main study parental consent form available in CD 1, Appendix 2) limited the total number of participants to 41: SI: 21, IF: 12, and C: 8. The reason behind the unequal number of participants in each of the 3 groups of the main study was due to the learners’ regular English class groupings, curricular constraints, parental consent requirements, and strict timetable limitations. The main study participants were also
not randomly assigned to the 3 experimental groups as they were grouped mostly according to their regular Greek class grouping. This also meant that they had not been uniformingly instructed in English.

Participant distributions in the pilot study and in the main study are presented in Table 1:

Table 1: Participant distributions

<table>
<thead>
<tr>
<th>Group</th>
<th>Pilot Study</th>
<th>Main Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI</td>
<td>11</td>
<td>21</td>
</tr>
<tr>
<td>IF</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>C</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>32</td>
<td>41</td>
</tr>
</tbody>
</table>

The participants in both the pilot study and the main study had already attended 3 years of English classes in the 1st, 2nd, and 3rd grade (since the studies were carried near the end of the school year) of Primary School. Their instruction in English had included implicit techniques such as songs, rhymes, role-plays, and drawing, designed for young learners, according to the State-appointed portfolios of English for the 1st and 2nd grade of Primary School along with materials especially designed by the teachers of the 3rd Model Experimental State Primary School of Evosmos. This fact enhances the likelihood that the participants had already been exposed to the target form. However, their performance in the interpretation pre-test and the production pre-test showed that they had minimal or no knowledge of the target structure under investigation. In the absence of proficiency tests, however, it is impossible to specify the exact proficiency level of these learners.
4.6. Research Design

The research design of both the pilot study and the main study could be summarized as including 2 independent variables: (a) instruction and (b) knowledge. The former consisted of 3 dependent variables: SI, IF, and C. The latter consisted of 4 dependent variables: performance in the interpretation pre-test and the production pre-test, and performance in the interpretation post-test and the production post-test.

4.7. The Pilot Study

4.7.1. Pilot study: introduction

The pilot study was conducted at the end of May – beginning of June 2013. It was a preliminary attempt to examine the issues under investigation. Based on the experience of the pilot study, the main study was fine-tuned and a number of important alterations were made. However, despite inadequacies, the pilot study was a full-scale study and it incorporated many of the elements similar studies require.

4.7.2. Pilot study: tokens

In the pilot study the selection of the tokens was made based on the verbs that 4th grade pupils are taught in the 4th grade State-appointed coursebook: ‘Αγγλικά Δημοτικού’ (‘4th Grade English’, Bratsoli & Diamantidou, 2010). This coursebook comprises the compulsory course material for 4th grade classes of English in all
Primary Schools in Greece. The decision to use the tokens of the 4th grade coursebook was based (a) on the fact that the pilot study was carried out at the end of the school year and (b) the decision to use verbs the 3rd grade learners were not familiar with, i.e. verbs 1 grade ahead their current grade level. In other words, it was expected that the participants would have had minimal exposure to the target verbs / tokens. After ensuring that learners were familiar (minimally) with these verbs, i.e. they knew at least the meaning of the verbs, target verbs were selected from chapters 1-7 of the coursebook (the whole coursebook consists of 10 chapters).

Accordingly, 78 regular English Past Simple Tense –ed verbs in total were available ranging from easily depictable ones, like ‘to cook’, to those which presented medium difficulty in order to be depictable, like ‘to exchange’, and to yet others that were very difficult to depict, e.g. stative verbs like ‘to realize’. These 78 regular English Past Simple Tense –ed verbs formed a corpus from which selection was made. The list of these pilot study verbs is available in CD 1, Appendix 3.

4.7.3. Pilot study: materials and implementation

The following sub-sections describe in detail the materials used in the pilot study for the implementation of the instructional treatments and the assessment measures. A sharp contrast between the materials of the pilot study and the materials of the main study concerns the amount of input offered in each case, an issue that will be discussed in detail in a subsequent section. The most notable similarity between the 2 studies involved the interpretation tests which were, in the main study, almost identical to those of the pilot study; albeit fine-tuned. Finally, the production tests utilized in the pilot study were probably more suitable for older learners, a fact that
was taken into consideration in the design of the main study production tests which were much easier than those of the pilot study.

4.7.3.1. Pilot study: materials and implementation: structured input

SI, as operationalized in the pilot study, consisted of 2 types of activities: (a) 4 referential tasks and (b) 4 affective tasks. The task instructions and the labels were provided in Greek to ease comprehension on the part of the participants and save time that might have been allocated to explain the tasks, had the instructions been provided in English. The distractors in all the SI pilot study activities included verbs in English Present Simple Tense form.

Beginning with the referential activities, in the 1<sup>st</sup> in sequence the participants read 15 sentences (10 regular and 5 distractors) and were required to put a tick (a) under the label ΠΑΡΟΝ (PRESENT) to show that the action denoted by the verb ‘is happening now, this moment’, or (b) under the label ΠΑΡΕΛΘΟΝ (PAST) to show that the verbal action referred to ‘past time’. Example 1 is a sample item from the 1<sup>st</sup> pilot study SI activity:

Example 1:

ΠΑΡΟΝ          ΠΑΡΕΛΘΟΝ

The English teacher wanted to test his pupils.

The 2<sup>nd</sup> pilot study referential activity required participants to read 15 sentences (10 regular and 5 distractors) and put a tick (a) under the label ΣΩΣΤΟ (CORRECT) to show if the sentence denoted an action that had occurred last year, or
(b) under the label ΛΑΘΟΣ (WRONG) in case the sentence was wrong (i.e. the action was situated in ‘present time’). Target form verbs were underlined. Example 2 is a sample item from the 2nd pilot study SI activity:

Example 2:

ΣΩΣΤΟ       ΛΑΘΟΣ

We played all kinds of sports.

In the 3rd pilot study SI activity, which was similar to the 1st pilot study SI activity, the participants read 15 sentences (10 regular and 5 distractors). These 15 sentences formed a storyline but were presented as separate propositions. The participants were required to put a tick (a) under the label ΠΑΡΟΝ (PRESENT) to show that the action denoted by the verb ‘is happening now, this moment’, or (b) under the label ΠΑΡΕΛΘΟΝ (PAST) to show that the verbal action referred to ‘past time’. Example 3 is a sample item from the 3rd pilot study SI activity:

Example 3:

ΠΑΡΟΝ       ΠΑΡΕΛΘΟΝ

Nick called Mary on the phone.

The 4th pilot study referential activity, which was similar to the 2nd pilot study SI activity, required participants to read 15 sentences (10 regular and 5 distractors). These 15 sentences formed a storyline but were presented as separate propositions. The participants were required to put a tick (a) under the label ΣΩΣΤΟ (CORRECT)
to show if the sentence denoted an action that had occurred in the past, or (b) under the label ΛΑΘΟΣ (WRONG) in case the sentence was wrong (i.e. the action was situated in ‘present time’). Target form verbs were underlined. Example 4 is a sample item from the 4th pilot study SI activity:

Example 4:

ΣΩΣΤΟ   ΛΑΘΟΣ

I started on a journey to the castle of a beautiful princess.

In the case of the pilot study affective tasks, there were no distractors. Instead, after completing the tasks, participants - always in pairs -, were asked to silently compare their answers with those of their partner. The following tasks are all pilot study SI affective tasks.

The 5th pilot study SI activity required that participants put a tick (a) under the label ΤΟ ΕΚΑΝΑ ΚΑΙ ΕΓΩ (I ALSO DID IT) in case they had performed the same action as the sentence denoted last weekend, or (b) under the label ΔΕΝ ΤΟ ΕΚΑΝΑ ΕΓΩ (I DID NOT DO IT) in case they had not performed the same action as the sentence denoted last weekend. Participants silently read 10 sentences and completed the task in pairs. Within-pair silent comparisons followed. Example 5 is a sample item from the 5th pilot study SI activity:

Example 5:

ΤΟ ΕΚΑΝΑ ΚΑΙ ΕΓΩ   ΔΕΝ ΤΟ ΕΚΑΝΑ ΕΓΩ

I brushed my teeth.
In the 6\textsuperscript{th} pilot study SI activity the participants were asked to imagine that with their English class they had performed a number of tasks the previous month. They were instructed to put a tick (a) under the label ΜΟΥ ΑΡΕΣΕ (I LIKED IT) in case they had enjoyed the task, or (b) under the label ΔΕ ΜΟΥ ΑΡΕΣΕ (I DID NOT LIKE IT) in case they had disliked the task. Within-pair silent comparisons followed. Example 6 is a sample item from the 6\textsuperscript{th} pilot study SI activity:

Example 6:

\begin{verbatim}
ΜΟΥ ΑΡΕΣΕ ΔΕ ΜΟΥ ΑΡΕΣΕ
\end{verbatim}

We learned many new English words.

The 7\textsuperscript{th} pilot study SI activity, which was similar to the 5\textsuperscript{th} pilot study SI activity, required that participants put a tick (a) under the label ΤΟ ΕΚΑΝΑ ΚΑΙ ΕΓΩ (I ALSO DID IT) in case they had ever performed the same action as the sentence denoted, or (b) under the label ΔΕΝ ΤΟ ΕΚΑΝΑ ΕΓΩ (I DID NOT DO IT) in case they had never performed the same action as the sentence denoted. Participants silently read 10 sentences and completed the task in pairs. Within-pair silent comparisons followed. Example 7 is a sample item from the 7\textsuperscript{th} pilot study SI activity:

Example 7:

\begin{verbatim}
ΤΟ ΕΚΑΝΑ ΚΑΙ ΕΓΩ ΔΕΝ ΤΟ ΕΚΑΝΑ ΕΓΩ
\end{verbatim}

I played with a dog.
In the 8th pilot study SI activity, which was similar to the 6th pilot study SI activity, the participants were asked to imagine that they had performed a number of tasks the previous month. They were instructed to put a tick (a) under the label ΜΟΥ ΑΡΕΣΕ (I LIKED IT) in case they had enjoyed the task, or (b) under the label ΔΕ ΜΟΥ ΑΡΕΣΕ (I DID NOT LIKE IT) in case they had disliked the task. Within-pair silent comparisons followed. Example 8 is a sample item from the 8th pilot study SI activity:

Example 8:

ΜΟΥ ΑΡΕΣΕ   ΔΕ ΜΟΥ ΑΡΕΣΕ

I opened new presents.

The 8 pilot study SI activities are available in CD 1, Appendix 4.

4.7.3.2. Pilot study: materials and implementation: input flood

All 8 pilot study IF tasks were designed and implemented in an exactly identical manner, and they were of the same format; content only differed. Each task presented participants with an age-appropriate short story. Each IF story consisted of 10 underlined verbs (in the regular English Past Simple Tense -ed form) which were the focal points of equal in number syntactic propositions. All the participants were required to do was read silently the stories and answer in writing 8 questions that were irrelevant to the aims of the study. The requirement to answer the questions was incorporated so as to give a purpose to the completion of the task, i.e. to ensure that the participants would read the story with interest. It was expected that this would
guarantee exposure to the embedded target structure tokens. Example 9 includes sample sentences from the beginning of the 1st pilot study IF story:

Example 9:

9 year-old Yiannis brushed his teeth. He cleaned up his room. He washed his hands in the bathroom. The boy watered the plants in the garden with his father and baked bread with his mother.

As already mentioned, the participants were informed beforehand that their task was to read the stories in order to answer in writing 8 written questions. These questions made use mostly of the verb to be and referred to side details. Example 10 presents the first 2 questions provided after the 1st pilot study IF story:

Example 10:

Is the boy 7 years old?

Is the boy’s name Yiannis?

Questions requiring target structure processing were not employed, to avoid a possible bias in the results. To conclude, in the pilot study IF instructional treatment the only requirement was that participants read the 8 stories and answer the study-irrelevant questions. All 8 pilot study IF tasks and the questions are available in CD 1, Appendix 5.

4.7.3.3. Pilot study: materials and implementation: interpretation tests
In the pilot study the participants in the interpretation pre- and post-tests were given 1 A4 size sheet of the sentences with corresponding As and Bs to circle. The instructor / researcher displayed the corresponding photographs to the whole group (SI or IF) on different occasions. The interpretation test photographs were printed on A4 size sheets (in colour) and each pair of A and B photographs were glued on a single approximately A2 size cardboard.

The participants simply read silently and listened to the sentences (which were read aloud by the instructor / researcher) for each experimental item prior to responding. Both the pilot study interpretation pre-test and the interpretation post-test were identical: Having read and heard each one of the 15 sentences (10 in regular English Past Simple Tense -ed form and 5 distractors in English Present Simple Tense form), participants were required to choose between 2 photographs (A and B) the one that best matched the sentence they had just read and heard.

As in the main study interpretation pre- and post-tests, in the pilot study interpretation pre- and post-tests there were no time adverbials (i.e. yesterday, then, before, etc.) or other indication of time (i.e. clocks, lighting effects like light for day or darkness for night, calendars, etc.) in the sentences. No time markers or signs appeared in the photographs. The ultimate intention was that the main cue to link the regular English Past Simple Tense –ed in the sentences to the corresponding photograph (A or B) was the understanding of the sequential order of events that the 2 photographs portrayed. The pilot study interpretation pre- and post-tests are available in CD 1, Appendix 6.

4.7.3.4. Pilot study: materials and implementation: production tests
As for the pilot study production pre- and post-tests, the participants were originally intended to read a story and be assessed on their oral re-production / re-narration of the story on a 1 to 1 basis with the instructor / researcher recording the oral data. Then, participants would be assessed in the written mode as a group. A number of factors mainly linked to time limitations did not allow this.

Consequently, in the pilot study, the 8-year old participants read and listened to a story (which was read aloud by the instructor / researcher) accompanied by photographs displayed by the instructor on cardboard. The sentences also appeared in the form of captions below each corresponding photograph. The photographs were then withdrawn and the participants were given a time limit to reproduce individually the story in written mode. The results showed that for some of them having to reproduce the whole story was a burdensome task. In fact, some of them found it very difficult and produced only around 2 to 3 sentences / propositions out of a total of 10 sentences / propositions. Their collective output, though minimal in some cases, was rewarding, nonetheless, as it contained other data, such as grammar (especially articles and plurals), syntax, evidence of learnability sequences, and so on. The pilot study production pre- and post-tests are available in CD 1, Appendix 7.

4.7.4. Pilot study: research schedule

As aforementioned, in order to measure the effects of instruction, interpretation pre- and post-tests and production pre- and post-tests were administered. The pre-tests were given to the participants 1 day before the instructional treatment session and the post-tests 1 day after. The whole experiment,
i.e. pre-tests → instruction → post-tests, lasted approximately 3 90-minute sessions. In more detail, the entire pilot study experiment lasted 3 90-minute sessions for the SI and the IF groups (including instruction and assessment), and 2 90-minute sessions for the C group (including assessment only). The post-tests were administered to all 3 groups at the same time in the same room on the final day of the pilot study. The pupils who had not been granted parental consent, although present, did not participate in any way in the study. They were appointed activities irrelevant to the present study. The entire pilot study research schedule is presented in Figure 6:

**Figure 6: Pilot study research schedule**

![Research schedule diagram]

4.7.5. Pilot study: scoring

Only the pilot study interpretation pre-test and the interpretation post-test were both transcribed and hand-scored by the instructor / researcher. The total score for the interpretation tests was calculated based on the number of correct A or B choices. 1
point was given for each correct answer. A wrong answer or inability to give an 
answer received a 0 mark. A correct response was one where participants correctly 
matched the sentence they read to 1 of the 2 pictures they had looked at. The score 
range for the experimental items was from 0 to 10 points.

The preliminary analysis of the interpretation pre-test and post-test data sets 
revealed a 19% decline in the performance of the SI group, a 29% decline in the 
performance of the IF group, and a 27% decline in the performance of the C group. 
These pilot study findings were not further statistically analyzed.

The production pre-test and the production post-test were transcribed 
analytically but not processed statistically in any way. However, acquisition order 
sequences and a host of other data on structures such as articles, plurals, etc. were 
discovered during the transcription process. Error patterns were also evident in the 
data. These findings required a combination of quantitative and qualitative analyses 
but were not statistically analyzed in any way.

4.7.6. Pilot study: summary

All in all, the pilot study served as a precursor to the main study. On the one 
hand, it tentatively proved that the SI and IF instructional methodologies could be 
implemented in contexts involving young learners. On the other hand, it brought to 
the surface a number of inadequacies that were taken into account in the design of the 
main study. The relevant changes that were made in the design of the main study are 
reported in the following section.

4.8. Main Study Changes Based on the Pilot Study
4.8.1. Main study changes based on the pilot study: introduction

In this section, the changes that shaped the main study will be reported with explicit reference to the problems faced in the pilot study and the ensuing solutions in the main study. These changes were necessitated (a) partly due to the inadequacies in the research design of the pilot study and (b) the related attempt to improve the research design of the main study.

4.8.2. The changes

Most of the changes that were made in the main study based on the shortcomings of the pilot study were related with (a) the attempt to render the materials more age-appropriate and (b) the aim to make implementation more effective. Such and other changes are reported in the following paragraphs.

To begin with, the 1st change was related to the fact that in the main study the number of tokens was reduced from 78 to 55. This affected the design of the main study materials in a number of profound ways. This issue is described in detail in the relevant main study section, i.e. 4.9.2.

The limited number of available tokens affected the main study materials in another, 2nd way, as well. Most of the available verbs were quite easy to display photographically (e.g. ‘to play’). However, approximately 5 to 7 verbs posed difficulties in 2 ways: (a) incorporation in a story (sequence of events / actions) and (b) pictorial presentation (e.g. ‘to stay’, or ‘to need’). To tackle this problem, the researcher consulted a well-known English dictionary (Quirk, 1990). Having come
across about 5 to 10 entries for each verb, the selection of the appropriate entry was made on the basis of 3 interrelated criteria: (a) how easy it would be to incorporate it in the story and depict it in pictorial form, (b) prototypicality of sense, and (c) how likely the targeted age group were to be acquainted with these verbs. Example sentences which appeared in the dictionary were copied and reformulated to fit the research requirements. These sentences were also used in the design of the accompanying photographs. The participants, with few exceptions, who required a direct oral translation in Greek, found no trouble with these tokens. The use of the corresponding photographs further eased the learners’ task.

The 3rd alteration was linked with the fact that some of the pilot study materials were not very age-appropriate. For example, the pilot study instructional treatment materials consisted of typewritten handouts with no accompanying photographs. The instructional treatment materials in the main study were modified as follows: (a) the instructional treatment materials were coloured so as to suit the age of the students, (b) more age-appropriate photographic materials accompanied both the SI and the IF instructional treatment activities with young children as models and these photographs were distributed on coloured laminated A4 and A3 size sheets to the learners, (c) in the SI activities, participants were asked to tick boxes under headings such as ✓ (yes), ✗ (no), 😊 (I liked it) ☹, (I did not like it), ♻ (now, at present), 🗿 (in the past), and (d) the sentences became shorter and simpler with respect to their syntax.

As to the 4th difference between the pilot study and the main study, in the main study the content of the instructional treatment material was more age-appropriate, too. In particular, 4 broad areas of young children’s life were focused on:
HOME, FREE TIME, the ENVIRONMENT, and SCHOOL. These 4 broad topics were chosen because the available verbs could be organized around them. These topics became the basis for the creation of 4 stories for the main study IF materials. The same exact sentences of the IF materials were copied, jumbled, and appropriately organized and rearranged to create the 4 main study SI activities, 2 of which also included a number of distractors. In other words, in the main study the SI and IF group participants were exposed to both the same tokens and the same exact sentences, whereas in the pilot study the SI and IF group participants were exposed to the same tokens, but not to the same exact sentences.

As to the 5th change, in the main study the size of the font was enlarged (each SI activity / IF story became 1 page long) and each activity was accessible visually on coloured laminated A4 size sheets distributed to individual participants. The aim of this change was to render the materials more attractive and suitable for the participants.

The 6th change involved another issue that was taken into account and aided in shaping the design and materials of the main study, i.e. the issue of the amount of time to be allocated in order to carry out the research schedule. Curricular constraints required that the whole procedure (pre-tests, instructional treatments, post-tests) be short and effective. 2 class hours (2 x 45 minutes = 90 minutes) were allocated to each of these 3 parts. One of the alterations that the pilot study dictated was that there should be ample time to perform the parts. Having taken into account a host of other factors, such as curriculum, participant age, number of available tokens, materials, desirable exposure to target form, participant attention span, pilot study findings etc., the aforementioned timetable was implemented. The fact that exposure time might be considered limited was not overlooked. However, it must be explicitly mentioned that
a longer exposure time span might have had the unwelcome effect that all participants would move to correct and accurate mastery of the target structure. Even though this is certainly welcome pedagogically, in studies like the present research project, a moderate amount of time is more appropriate in order to validly assess the effects of the instructional treatments involved. The necessary time was calculated accurately in order to ensure that high-performing participants would not be distracted and low-achievers would not find the task insurmountable.

The 7th change, or, to be more accurate, methodological decision concerned the mode, oral or written, that participants were required to respond in. As regards the instructional treatments, the original plan was to include an equal number of activities in both modes: oral and written. Based on the execution of the pilot study, in the main study all the SI group and IF group instructional treatment activities required participants only to read silently and respond. In the testing measures (interpretation and production), participants simply read silently and listened to the materials which were read aloud by the instructor before responding.

The decision to read aloud the testing materials was taken on the spot by the instructor / researcher. While carrying out the interpretation pre-test task, the instructor / researcher discovered (both by direct observation and by relevant comments by 2 or 3 participants) that the data load for these very young children was still burdensome. As a result, the instructor / researcher decided to ease the task by actually reading aloud both the sentences (once, in the interpretation tests) as well as the stories (twice, in the production tests).

As to the 8th change, this concerned the (re)design of the main study assessment measures: (a) the interpretation pre-test and post-test, and (b) the production pre-test and post-test, based on the problems faced in the pilot study.
Beginning with the interpretation tests, both the pre-test and the post-test were of the same format in both studies. As aforementioned, in both the pilot study and the main study, having read 15 sentences (10 in regular English Past Simple Tense -ed form and 5 distractors in the English Present Simple Tense form), participants were required to choose between 2 corresponding photographs (A or B) the one that best matched the sentence they had read and heard. Up to this point, the pilot study and the main study interpretation tests were identical. Nonetheless, the pilot study brought to the surface a number of interrelated issues that affected the main study interpretation tests as well as other parts of the research project: (a) ease of carrying out the experiment on the part of the instructor / researcher, (b) allocation of an exactly similar amount of time for the participants to study the material, (c) visibility of the material on the part of the participants, (d) classroom seating arrangements, and (e) the need to minimize any kind of unnecessary instructor / researcher interference.

To be more specific, in the pilot study the participants in the interpretation tests were given 1 A4 size sheet of the sentences with corresponding As and Bs to circle. The instructor / researcher displayed the corresponding photographs to each group (SI, IF, or C) at the same time for the whole group on different occasions. The interpretation test photographs (in colour) in the pilot study were printed on A4 size pages and each pair of A and B photographs were glued on a single approximately A2 size cardboard. However, this resulted in making the whole set heavy to carry, oversized, and difficult to administer. What is more, it became obvious that a few participants did not have enough time to study the photographs which were presented once for about 10 seconds and never appeared again. A few other participants could not easily make out the photographic material. The regular classroom seating arrangements had not been taken into account, i.e. children were seated / arranged in
small groups usually consisting of up to 6 children (3 desks / pairs) and did not always have direct physical eye contact with the whiteboard in front of which the photographs were demonstrated by the instructor / researcher. Finally, stress and anxiety factors, which might have also affected the time allocated to exhibiting each set of photographs, on the part of the instructor / researcher, brought to light the fact that his involvement should be minimized if not zeroed.

The solution to all these problems was to reshape the design of the photographic materials in a way that made the implementation of the materials simple, easy, speedy, and accurate. Consequently, in the main study the photographs (in colour) were re-shot according to the new tokens available, copied and pasted on A4 size computer sheets, minimized to approximately 10% of the original size, printed on A3 size sheets which were then laminated. Therefore, in the main study the material for each interpretation test was 1 A4 size copy of the sentences with corresponding As and Bs to be circled (identical to the pilot study) and 4 A3 size laminated sheets (consisting of 2 A4 size sheets) each consisting of approximately 8 double photographs (8 X 2 = 16 photographs). Each pair of participants, who were usually seated in pairs in 1 desk, received 1 A4 size copy of the sentences with the As and Bs as well as 4 sets of photographs on A3 size sheets. This guaranteed that an exactly equal amount of time was to be allocated for the pre- and post-tests and that all the instructor / researcher had to do was distribute the material and ensure that the prearranged time limit was adhered to.

Moreover, the random selection of photographs to be presented as A or B in the pilot study was fine-tuned in the main study so that each interpretation test (pre-test and post-test) consisted of an equal amount of A and B photographs as the correct response item. Therefore, in the main study interpretation pre-test, there were 5
correct A responses and 5 correct B responses for the regular tokens, and 3 correct A responses and 2 correct B responses for the distractors. A similar design characterized the post-test with the only difference that there were 2 correct A responses and 3 correct B responses for the distractors. Meticulous attention was also given to differentially and randomly organizing the order of As and Bs as the correct answer in the interpretation pre- and post-tests.

As to the production pre- and post-tests, in the pilot study, the 8-year-old participants read and listened to a story (the story was narrated by the instructor / researcher) accompanied by photographs on cardboard (with the corresponding sentences below the photographs) and were given a time limit to reproduce the story in written mode collectively. The findings showed that for some of them having to reproduce the whole story was a burdensome task. In order to overcome this overload and focus exclusively on the target form, in the main study participants read and listened to the story (the story was narrated by the instructor / researcher) accompanied by the photographs (all in new format) and were then handed over the story with gaps requiring them only to write down the missing verbs in the regular Past Simple Tense -ed form (10 tokens, all obligatory contexts). This made implementation easier, output more uniform and more easily assessable, and analysis more accurate.

The 9th change was related with the fact that the amount of time the participants in the instructional treatment conditions spent on their tasks was more accurately balanced in the main study. More specifically, in the pilot study the participants received the SI tasks in the form of a handout including all the tasks. The instructor explained all the tasks in the handout at once, pausing after each task for queries, and withdrew. This had as a result that some learners finished early and
simply waited, whereas others finished later. Consequently, individual learners spent an unequal amount of time on the SI material. A similar phenomenon arose during the implementation of the IF materials. More specifically, in the pilot study the participants received the IF tasks in the form of a handout including all the stories. Their task was the same for all stories, i.e. to read silently the stories and answer in writing the study-irrelevant questions. This had the unwelcome effect that individual participants spent unequal amounts of time to complete the IF tasks, as well. In fact, some participants finished much earlier than expected, whereas others completed their task much later. Despite the fact that the early-finishers sat silently in their seats, some late-finishers started asking questions as the tasks were obviously difficult for them.

In order to tackle the issue of amount of time spent on the tasks, a different procedure was adopted in the main study. First, each SI and IF task, along with the photographic material, was distributed one at a time. Then, for the SI group, the instructor explained each task after the participants had completed the previous one. For the IF group, the necessary instructions to implement the stories were provided once at the beginning. When the participants had completed each task, the materials and the photographs were withdrawn and the next task was carried out. Despite the fact that the exact amount of time each individual participant spent on each task was not accurately balanced, even in this way, this procedure was more accurate than the one adopted in the pilot study. In the main study the instructor / researcher timed all the tasks the participants carried out and found out that the total amount of time the SI and IF group participants spent on all the respective tasks was exactly the same, i.e. 58 minutes.

The fact that the materials were distributed, instructions were given, and then the materials were withdrawn, for each SI and IF activity, proved time-consuming.
This can partly explain why out of a 90-minute session, i.e. the approximately total amount of time the instructor was in the classroom during the instructional treatments, actual participant exposure to the materials lasted only 58 minutes. The remaining time was spent waiting for learners to enter the classroom after the morning prayer or the break, welcoming learners, waiting for them to be seated, ensuring the classroom atmosphere was appropriate to commence the experiment, distributing the material to individual participants, providing task instructions, collecting the materials after the completion of each task, and, after the instructional treatment was over, thanking learners and the class teachers.

Finally, ease of implementation led to the abandonment of another idea that had initially seemed very age-appropriate. More specifically, in order to make the materials more game-like, it was proposed that in the main study SI treatment, participants, instead of answering, would stick relevant stickers personally designed and created by the instructor / researcher. The idea came up when the instructor / researcher was trying to find a symbol to denote \textit{past time}. An anonymous informant (unaware of the present research agenda) told him that in Sign Language the notion of \textit{past} is denoted with something like a movement of one’s hand above the right shoulder (thumb pointed) and \textit{present time / now} is depicted with an up and down pointing of the right hand (pointer pointed). A search for similar symbols that would universally and undeniably represent \textit{present} and \textit{past} ensued. The symbols were found in a different format in the Symbol section of MsWord. They were transformed into Adobe Acrobat format, rearranged, and redesigned. \begin{lstlisting}[language=TeX] became a symbol to denote \textit{past} and \begin{lstlisting}[language=TeX] was used to refer to \textit{present}. The symbols were then copied to MsExcel and coloured, and, alongside other symbols, were subsequently printed on
ready-made sticker pages available in bookshops. The result was indeed rewarding but would have made implementation more difficult and distracting for such young learners. The symbols, however, were maintained and the obstacle of pictorially portraying present time and past time successfully overcome. The original sticker page which was not used is available in CD 1 Appendix 8.

4.8.3. Main study changes based on the pilot study: summary

The implementation of the pilot study brought to the surface a number of limitations in the research design and execution of the pilot study experiment that aided in the design of the main study. The methodological decision of using bolding in the IF materials in the main study, instead of underlining in the pilot study, should be highlighted. A number of changes involved the execution of the experiment on the part of the instructor / researcher. These focused mostly on minimizing his role. This was indeed achieved as a number of on-the-spot decisions in the main study were made possible. Also, this was a desired outcome, since it meant that the results were probably only minimally affected by the presence of the instructor / researcher since his major roles were to distribute the materials, explain the instructions, solve unexpected problems, and be wholesomely available for unexpected occurrences. The major contribution of the pilot study was the conclusion that the SI and IF & TIE instructional treatments were indeed adaptable to young 8-year-old Greek L1 EFL learners, a target group rarely experimented on in the relevant literature, especially within the SI framework.

4.9. The Main study
4.9.1. Main study: introduction

Under the prism of the aforementioned alterations, a fine-tuned full-scale study, i.e. the main study, was carried out at the end of May 2014, near the end of the school year in Greece, where lessons in Primary Schools officially end on the 15th of June. Even though the overall results of the data analyses of the performance of the 3rd graders in the main study cannot unequivocally constitute evidence of acquisition of the target structure, nonetheless, pre-test to post-test performances were statistically significant in a number of cases, a fact which tentatively leads to the conclusion that the instructional treatments did have an effect on learners’ performance. A major contribution of the main study is the suggestion that, as to EFL teaching, the most important concern should not involve content, e.g. targeted grammatical structures such as the regular English Past Simple Tense –ed, but, instead, research and classroom practice should focus more on making the teaching methodologies, techniques, and materials more adaptable to and suitable for the learners.

4.9.2. Main study: tokens

The main study was conducted based on a different coursebook, i.e. the 3rd grade Primary School English coursebook, targeting learners of lower EFL proficiency, than the 4th-grade coursebook utilized in the pilot study. There are 2 books, i.e. Magic Book 1 and Magic Book 2, for 3rd graders in all Greek Primary Schools. The former addresses complete beginners in English, while the latter addresses those learners who started learning English from grade 1. Magic Book 1
(Alexiou and Mattheoudakis, 2012) was used for the selection of tokens. As the number of tokens was limited to 47, the number of activities, for both the SI and the IF instructional treatment packets, was limited from 8 in the pilot study to 4 in the main study and the majority of tokens were used twice in the main study: in the assessment measures and in the instructional treatments. A number of regular English Past Simple Tense –ed verbs not present in the Magic Book 1 coursebook were also included since some of the available verbs were not considered easily depictable. This increased the total number of available verbs to 55. The participants exhibited knowledge of the verbs and there was no need for a translation into Greek, especially taking into account that these verbs were accompanied by relevant photographs (e.g. ‘to phone’, or ‘to rain’). The complete list of the main study verbs is available in CD 1, Appendix 9.

4.9.3. Main study: materials and implementation

In the following subsections, the materials of the SI and IF & TIE instructional treatments, as well as the interpretation and production tests utilized in the main study will be described. References to the pilot study will not be often reiterated, even though in a number of cases the presentation will allude to information from previous sections and subsections describing the pilot study.

4.9.3.1. Main study: materials and implementation: structured input

In actual practice, SI, as operationalized in the main study, consisted of 2 types of activities: (a) 2 referential tasks, and (b) 2 affective tasks. The task instructions and
the labels were provided in Greek to ease comprehension on the part of the participants and save time that might have been allocated to explain the tasks, had the instructions been provided in English. The distractors in all the SI main study activities included verbs in English Present Simple Tense form.

Beginning with the referential activities, the 1st in sequence was related to the topic HOME. The participants read 15 sentences (10 regular and 5 distractors) accompanied by corresponding photographs and had to tick (a) a box under the symbol \( \text{not happening now, this moment} \) to show that the action denoted by the verb ‘is happening now, this moment’, or (b) a box under the symbol \( \text{past time} \) to show that the verbal action referred to ‘past time’. After briefly reading aloud the instructions, the instructor completely withdrew. The participants were left to read silently as they completed the task aided by the photographic material. Example 11 is a sample item from the 1st main study SI activity:

**Example 11:**

I cooked spaghetti. (photo 1)  
\[ \square \]  
\[ \square \]

The 2nd main study activity, loosely organized around the theme FREE TIME, required students to read 15 sentences (10 regular and 5 distractors) and tick (a) a box under the label \( \text{last year (correct)} \) to show if the sentence denoted an action that had occurred last year (correct), or (b) a box under the label \( \text{incorrect} \) in case the sentence was wrong (i.e. the action was situated in the present). After briefly reading aloud the instructions, the
instructor completely withdrew. The participants were left to read silently as they completed the task aided by the photographic materials. Example 12 is a sample item from the 2nd main study SI activity:

Example 12:

They loved the playground. (photo 2)  

In the case of the affective tasks, there were no distractors. Instead, after completing the tasks, participants - always in pairs - were asked silently to compare their answers with those of their partner. ENVIRONMENT was the thematic unit around which the 3rd main study SI activity (1st main study affective activity) was loosely organized. The instructor read out the rubric, instructing the participants to tick the appropriate boxes under the relevant labels in case they had done the same thing as the sentence denoted in the past (✓) or not (✗). Participants silently read 10 sentences, which were accompanied by corresponding photographs, and completed the task in pairs. Within-pair silent comparisons followed. Example 13 is a sample item from the 3rd main study SI activity:

Example 13:

We recycled. (photo 9)
Likewise, under the topic SCHOOL, the children carried out the 2\textsuperscript{nd} affective activity, the 4\textsuperscript{th} and final one of the main study SI intervention packet. In this case, they were asked to imagine that with their English class they had performed a number of tasks the previous months. They ticked a box under the happy face (☺), in case they had enjoyed the task, or a box under the sad face (☹), in case they had disliked the task. Participants silently read 10 sentences, which were accompanied by corresponding photographs, and completed the task in pairs. Within-pair silent comparisons followed. Example 14 is a sample item from the 4\textsuperscript{th} main study SI activity:

Example 14:

\[
\begin{array}{c}
\text{☺} \\
\begin{tabular}{l}
We traced the letters. (photo 1) \\
\hline
\end{tabular}
\end{array}
\]

All 4 main study SI tasks are available in CD 1, Appendices 10A, 10B, 10C, and 10D. The corresponding photographs are available in CD 2, Appendices 1A, 1B, 1C, and 1D.

4.9.3.2. Main study: materials and implementation: input flood

All 4 main study IF tasks were operationalized in an exactly identical manner and they were of the same format; content only differed. Based on the same topics as the main study SI tasks (also in their order of presentation), each task presented participants with an age-appropriate short story. In order to accurately balance the amount of exposure to the target structure in the 2 instructional packets, the IF
sentences were used intact to form the SI tasks; the only difference was that in the SI
tasks these sentences did not form a chronologically evolving storyline but were
deliberately presented in random order. The exact same photographs were used in
both instructional treatments with the provision of 10 additional photographs required
for the distractors in the 1st and 2nd main study SI (referential) activities. Each main
study IF story consisted of 10 bolded verbs (in the regular English Past Simple Tense
-ed form) which were the focal points of equal in number syntactic propositions. All
the participants were required to do was read silently the stories and study the relevant
photographs at their ease. Example 15 includes sample sentences from the beginning
of the 1st main study IF story:

Example 15:

I opened my eyes (photo 1)! In the bathroom, I brushed my teeth (photo 2)
and washed my face (photo 3).

Participants were informed beforehand that their task was to read the stories in
order to answer orally a series of oral questions asked by the instructor / researcher.
The requirement to answer the questions was incorporated so as to give a purpose to
the completion of the task, i.e. to ensure that the participants would read the story with
interest. These questions made use mostly of the verb to be and referred to side details
such as presented in Example 16:

Example 16:

What colour is the broom?
Where are the girls?

Questions requiring target structure processing were not employed, to avoid a possible bias in the results. To conclude, in the main study IF instructional treatment the only requirement was that participants read the 4 stories, study the corresponding photographs, and answer the study-irrelevant questions. All 4 main study IF tasks are available in CD 1, Appendices 11A, 11B, 11C, and 11D. The IF questions are available in CD1, Appendix 11E. The corresponding photographs are available in CD 2, Appendices 2A, 2B, 2C, and 2D.

4.9.3.3. Main study: materials and implementation: interpretation tests

The implementation of the main study interpretation tests was in the following form: The participants read 10 sentences and also listened to the instructor / researcher read aloud once these sentences including regular English Past Simple Tense -ed verbs, accompanied by an additional 5 distractor sentences (in English Present Simple Tense form) and chose between 2 pictures (A or B) the one that correctly matched the sentence before them. Each pair of participants was given by the instructor 4 A3 size laminated sheets which included all the double photographs numbered from 1 - 15 (interpretation pre-test) and 16 - 30 (interpretation post-test). Their task now was to write their personal unique number, read the sentences read aloud by the instructor / researcher, study the corresponding photographs, and circle the corresponding As or Bs. As in the pilot study, there were no time adverbials (i.e. yesterday, then, before, etc.) or other indication of time (i.e. clocks, lighting effects like light for day or darkness for night, calendars, etc.) in the sentences. No time
markers or signs appeared in the photographs. The ultimate intention was that the main cue to link the regular English Past Simple Tense –*ed* in the sentences to the corresponding photograph (A or B) was the understanding of the sequential order of events that the 2 photographs portrayed. The instructions were given in Greek to avoid unnecessary confusion. The main study interpretation pre- and post-tests are available in CD 1, Appendices 12A and 12B. The corresponding photographs are available in CD 2, Appendices 3A and 3B.

**4.9.3.4. Main study: materials and implementation: production tests**

In the main study the participants read and listened to a story accompanied by photographs (all in new format) and were then handed over the story with gaps requiring them to fill in the missing verbs in the regular English Past Simple Tense –*ed* form (10 tokens, all obligatory contexts). The materials for the production pre-test included 1 A4 size laminated copy of the stories without gaps and another A4 size copy of the stories with 10 gaps, 1 gap for each token. The materials for the production post-test were also 1 A4 size laminated copy of the stories without gaps and another A4 size copy of the stories with 10 gaps, 1 gap for each token.

The implementation of the main study production tests was in the following form: The subjects first read silently on their own a story containing 10 short and simple propositions, accompanied by 10 photographs (on 1 A4 size laminated sheet) depicting the actions / verbs (10 regular Past Simple Tense -*ed* verbs). Then, the story was read twice by the instructor / researcher and studied by the children for 10 minutes before both the complete story and the accompanying photographs were withdrawn. The children were then given another identical A4 size page with the
verbs missing. The participants had to complete the task by filling in the correct verb in English Past Simple Tense –ed form. As said, each verbal action was depicted in photographic form. To avoid unnecessary confusion, a parenthesis with the corresponding photograph number (i.e. photo 7, photo 8, etc.) was included after each verb or gap. The main study production pre- and post-tests are available in CD 1, Appendices 13A and 13B. The versions of the production pre- and post-tests without the gaps are available in CD1, Appendices 13C and 13D. The corresponding photographs are available in CD 2, Appendices 4A and 4B.

4.9.4. Main study: research schedule

As aforementioned, in order to measure the effects of instruction in the main study, interpretation and production pre-tests and post-tests were administered. The pre-tests were given to the participants 1 day before the instructional treatment session and the post-tests 1 day after. The entire experiment, i.e. pre-tests → instruction → post-tests, lasted approximately 3 90-minute sessions. In more detail, the entire main study experiment lasted 3 90-minute sessions for the SI and the IF groups (including instruction and assessment), and 2 90-minute sessions for the C group (including assessment only). The SI and the IF groups received instructional treatment interventions on different occasions at the same day. The post-tests were administered on a separate occasion for the SI group, and on a separate occasion at the same day at the same time for the IF and C groups in the same room on the 3rd and final day of the main study experiment. The pupils who had not been granted parental consent, although present, did not participate in any way in the study. They were appointed
activities irrelevant to the present study. The entire main study research schedule, identical to that of the pilot study, is presented in Figure 7:

**Figure 7: Main study research schedule**

```
Instruction

90 minutes

Pre-tests
90 minutes

Post-tests
90 minutes
```

The timing of the instructional treatments is presented in Table 2:

**Table 2: Timing of main study instructional treatment tasks**

<table>
<thead>
<tr>
<th></th>
<th>SI</th>
<th>IF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; Task</td>
<td>08:25 - 08:42</td>
<td>12:45-13:02</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; Task</td>
<td>08:50 - 09:05</td>
<td>13:03-13:15</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; Task</td>
<td>09:10 - 09:25</td>
<td></td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt; Task</td>
<td>09:28 - 09:38</td>
<td></td>
</tr>
</tbody>
</table>

Total: **58 minutes**
The total amount of time required by the participants to complete the tasks in the main study was identical for both the SI and the IF groups: 58 minutes.

The time allocated for the tests is presented in Table 3:

<table>
<thead>
<tr>
<th>Task</th>
<th>Time</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd Task</td>
<td>13:22-13:41</td>
<td>19 minutes</td>
</tr>
<tr>
<td>4th Task</td>
<td>13:43-13:53</td>
<td>10 minutes</td>
</tr>
</tbody>
</table>

**Total: 58 minutes**

Table 3: Duration of main study assessment measures

<table>
<thead>
<tr>
<th>Assessment Measure</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpretation pre-test</td>
<td>27 minutes</td>
</tr>
<tr>
<td>Interpretation post-test</td>
<td>22 minutes</td>
</tr>
<tr>
<td>Production pre-test</td>
<td>35 minutes</td>
</tr>
<tr>
<td>Production post-test</td>
<td>30 minutes</td>
</tr>
</tbody>
</table>

The reduced amount of time in the post-tests can be explained by test familiarity effects, as the pre-tests and the post-tests were exactly identical in format, only the tokens and the corresponding photographs differed.

**4.9.4. Main study: scoring**

All the main study assessment measures, i.e. the interpretation pre-test and post-test, and the production pre-test and post-test, were transcribed and hand-scored by the instructor/researcher. The total score for the interpretation test was calculated based on the number of correct A or B choices. 1 point was given for each correct answer. A wrong answer or inability to give an answer received a 0 mark. A correct response was one where participants correctly matched the sentence they read and
heard to 1 of the 2 pictures they looked at. The score range for the experimental items was from 0 to 10 points. The answers to the distractors in the interpretation tests were saved on a separate file but were not statistically analyzed.

The preliminary analysis of the main study interpretation pre-test and post-test data sets revealed a 13% decline in the performance of the SI group, a plateau (0% difference between pre-test and post-test) in the performance of the IF group, and a 6% increase in the performance of the C group. The findings were further statistically analyzed and the results are presented in chapter 5.

Based on the experience drawn from the results of the pilot study, scoring for the production tests was simplified and rendered straightforward in the main study. The participants’ task in the main study was to supply the correct verb in the regular English Past Simple Tense -ed. A 0 to 1 scoring system was used. 1 point was assigned to exactly correct responses (i.e. ‘loved’) or responses that were orthographically wrong but phonologically correct (i.e. ‘lovd’ or ‘lisent’). The term correct is used to refer to any answer that presented knowledge, acquisition / learning, or simple provision of the respective sound of –ed in written form as it is pronounced in the different contexts. All other responses, i.e. Present Simple Tense forms (e.g. ‘plays’), bare infinitives (e.g. ‘correct’), wrong verb use (e.g. ‘stayed’ instead of ‘loved’), and no answer, received a 0 mark. Again, the score range was from 0 to 10 points. There were no distractors in the production tests.

The preliminary analysis of the main study production pre-test and post-test data sets revealed a 26% increase in the performance of the SI group, a 30% increase in the performance of the IF group, and a 30% increase in the performance of the C group. The findings were further statistically analyzed and the results are presented in chapter 6.
4.9.5. Main study: summary

Despite the fact that the increase as to the performance of the C group in the main study points to possible inadequacies, the design of the materials based on the experience of the pilot study minimizes the role of such inadequacies which undoubtedly exist. A major contribution of the main study was that, in one way or another, the instructional treatments did have an effect on participants’ performance. The detailed data analyses that were performed and are presented in chapters 5 and 6 revealed a host of other interesting findings, as well.

Conclusion

As a concluding remark, it needs to be underscored that the attempt to answer the research questions, i.e. whether the SI and IF & TIE implicit methodologies are appropriate for the acquisition of the regular English Past Simple Tense –ed target structure (research question 1) and whether the SI and IF & TIE implicit methodologies are appropriate for the acquisition of the target structure by young Greek L1 EFL learners (research question 2), lead to the execution of 2 full-scale studies, a pilot study and the main study. Although the SI and IF methodologies, and the TIE techniques are not novel within the SLA research framework, both studies adapted the materials of such methodologies and techniques to suit the needs of young non-Romance Greek L1 EFL learners. This applies especially for the SI research paradigm where most studies have included mostly adult Spanish L1 EFL or ESL learners.
Despite the fact that the results of the data analyses of the main study are reported in the following 2 chapters, as a preliminary remark, the statistical analyses proved that instruction as operationalized in the main study did exhibit signs of acquisition of the regular English Past Simple Tense –ed on the part of the learners. What is more, the data analyses revealed a number of unexpected results that on the surface discredit the research design utilized, but might also be interpreted as signs that young learners respond to such instructional treatments differently than adults.

As to the research hypotheses, the fact that they were partly confirmed does not in any way discredit the methodological implementations adopted in the main study. This state of affairs may also be attributed to the unique traits of the target group as well.
CHAPTER 5:

RESULTS: INTERPRETATION

Introduction

This chapter includes the results of the statistical analyses of the performance of the participants in the main study in the interpretation pre-test and in the interpretation post-test. The presentation of the results is divided in the following sections: In section 5.1., the whole sample is presented. In section 5.2., the mean performance values and the standard deviation values of the whole sample are included and statistically significant differences between the interpretation pre-test and the interpretation post-test data sets (including all 41 participants) are reported. In section 5.3., the mean performance values and the standard deviation values of each of the 3 groups (SI, IF, and C) separately are included and statistically significant differences of the 3 groups separately are reported. In section 5.4., the statistically significant differences in all data sets of the interpretation tests are summarized. In section 5.5., individual participants’ performance is analyzed. In section 5.6., an item analysis of the tokens in the pre-test and in the post-test is included. The chapter concludes with a short summary of the most important findings of the statistical analyses as to the interpretation of the target structure.

5.1. Sample Analysis
41 participants took part in the main study. There were 21 participants in the SI group, i.e. 51.2% of the whole sample, 12 participants in the IF group, i.e. 29.3% of the whole sample, and 8 participants in the C group, i.e. 19.5% of the whole sample. All the young learners who participated in all parts of the main study were included in the statistical analyses, i.e. there were no missing values.

5.2. Interpretation: Whole Sample

5.2.1. Interpretation: whole sample: mean performance & standard deviation values

Table 4 presents the mean performance values and the standard deviation values of the variables interpretation pre-test and interpretation post-test for the whole sample of the 41 participants in the main study.

Table 4: Interpretation pre-test and post-test mean performance values and standard deviation valuea (whole sample)

<table>
<thead>
<tr>
<th>MEASURE</th>
<th>INTERPRETATION PRE-TEST</th>
<th>INTERPRETATION POST-TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cases</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>Mean</td>
<td>4.2</td>
<td>3.7</td>
</tr>
<tr>
<td>Std</td>
<td>1.3</td>
<td>1.6</td>
</tr>
</tbody>
</table>
According to Table 4, the interpretation pre-test mean performance value of the whole sample was 4.2 points and the interpretation post-test mean performance value was 3.7 points exhibiting a drop of approximately 0.5 points between the interpretation pre-test and the interpretation post-test. The interpretation pre-test standard deviation value for the whole sample was 1.3 and the interpretation post-test standard deviation value was 1.6 exhibiting an increase between the interpretation pre-test and the interpretation post-test.

5.2.2. Interpretation: whole sample: statistically significant differences

The performance of the whole sample of 41 learners in the interpretation pre-test and the interpretation post-test are presented in Histograms 1 and 2 respectively:

**Histogram 1: Interpretation pre-test performance frequencies for the SI, IF, and C groups combined (normal distribution bell curve line included)**
Histogram 2: Interpretation post-test performance frequencies for the SI, IF, and C groups combined (normal distribution bell curve line included)
According to Histogram 1 (interpretation pre-test), either the sample is indeed representative of the whole (possible) population of learners of this age exposed to the target structure when it comes to learning capacity, and / or the interpretation pre-test assessment tool used in the present study was efficiently designed to procure a normally distributed performance by the 41 participants. The interpretation pre-test was neither too difficult, nor too easy for these participants.

In Histogram 2 (interpretation post-test), the most notable observation is that there is a clustering of performance values in the 3 points value (16 cases) obviously digressing from the normal distribution bell curve line included. This could probably mean, either that the sample is not representative of the whole (possible) population of
learners of this age exposed to a second language grammatical form when it comes to learning capacity, and/or that the interpretation post-test assessment tool used in the present study was not efficiently designed to procure a normally distributed performance by the 41 participants. The interpretation post-test seems to have been difficult for these participants.

Obviously, there was less variation in the interpretation pre-test (Histogram 1) than in the interpretation post-test (Histogram 2). In order, however, to accurately verify or not whether the values of a data set represent a normally distributed sample, i.e. in order to verify or not the null hypothesis, graphic representations are not sufficient. They do offer initial visual information but a standard statistical analysis is also required. There are a number of relevant statistical analyses available, but the most powerful is the Shapiro-Wilk normality test.

Consequently, the interpretation pre-test data set and the interpretation post-test data sets (SI, IF, and C groups combined) were submitted to the Shapiro-Wilk statistical analysis normality test and the results are presented in Table 5:

<table>
<thead>
<tr>
<th></th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
</tr>
<tr>
<td>Pre-test</td>
<td>.952</td>
</tr>
<tr>
<td>Post-test</td>
<td>.829</td>
</tr>
</tbody>
</table>

The results of the Shapiro-Wilk statistical analysis can be explained as indicating a different distribution pattern for the interpretation pre-test and the
interpretation post-test. More specifically, for the interpretation pre-test the value of significance was p=0.084. Consequently, the null hypothesis, i.e. that the interpretation pre-test data set constituted a sample exhibiting a normal distribution pattern, was confirmed. For the interpretation post-test the significance value was p=0.00. Consequently, the null hypothesis, i.e. that the interpretation post-test data set constituted a sample exhibiting a normal distribution pattern, was rejected.

The final step was to examine whether there was a statistically significant difference as to the performance of the whole sample of 41 participants (i.e. of the SI, IF, and C groups combined) between the interpretation pre-test and the interpretation post-test. Since the whole sample of the 41 participants’ interpretation pre-test data set exhibited a normal distribution pattern, whereas the whole sample of the 41 participants’ post-test data set did not exhibit a normal distribution pattern, a non-parametric statistical analysis, i.e. the Wilcoxon Signed Ranks Test, was performed. The results are presented in Table 6:

Table 6: Interpretation post-test to pre-test statistical analyses (whole sample)

<table>
<thead>
<tr>
<th>Group</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>-2.070</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>0.038</td>
</tr>
</tbody>
</table>

As aforementioned, the interpretation pre-test mean performance value for the whole sample was 4.2 points and the interpretation post-test mean performance value was 3.7 points. The significance value produced was p=0.038. In other words, the
difference as to the performance of the whole sample of 41 learners who participated in the present study between the interpretation pre-test and the interpretation post-test reached a level of statistical significance.

5.2.3. Interpretation: whole sample: summary

The findings in this section can be summarized along the following lines. First, the mean performance of all 41 participants decreased between the interpretation pre-test and the interpretation post-test by 0.5 points and this decrease reached a level of statistical significance (p=0.038). Second, the standard deviation value was higher in the post-test (1.6) than in the pre-test (1.3). Finally, according to the results of the normality tests, the pre-test data set exhibited a normal distribution pattern (p=0.084), whereas the post-test data set did not (p=0.00).

5.3. Interpretation: SI, IF, and C groups

5.3.1. Interpretation: SI, IF, and C groups: mean performance & standard deviation values

Figure 8 presents the mean performance values as to the interpretation of the target structure of the 2 experimental groups of learners that participated in the main study (SI, IF) and the control group (C) both before the administration of the experimental procedures (interpretation pre-test) and after the administration of the experimental procedures (interpretation post-test).
According to Figure 8, after being exposed to the SI experimental condition, the SI group learners’ mean performance as to the interpretation of the target structure dropped by 1.3 points. After being exposed to the IF experimental condition, the IF group learners’ mean performance as to the interpretation of the target structure remained unaffected (0 points difference). The C group of learners’ mean performance as to the interpretation of the target structure increased by 0.6 points.
The interpretation pre-test and the interpretation post-test data sets of each of the 3 groups of learners (SI, IF, and C) separately were further analysed and the standard deviation values produced are presented in Table 7:

Table 7: SI, IF, and C group interpretation pre-test and post-test standard deviation values

<table>
<thead>
<tr>
<th>MEASURE</th>
<th>SI Pre-test</th>
<th>SI Post-test</th>
<th>IF Pre-test</th>
<th>IF Post-test</th>
<th>C Pre-test</th>
<th>C Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std</td>
<td>1.1</td>
<td>0.9</td>
<td>1.4</td>
<td>2.3</td>
<td>1.6</td>
<td>1.8</td>
</tr>
</tbody>
</table>

According to Table 6, in the interpretation pre-test the standard deviation value for the C group of learners was 1.6. The pre-test standard deviation value for the IF group of learners was 1.4. The pre-test standard deviation value for the SI group of learners was 1.1. In the interpretation post-test, the standard deviation value for the IF group of learners was 2.3. The post-test standard deviation value for the C group of learners was 1.8. The post-test standard deviation value for the SI group of learners was 0.9.

5.3.2. Interpretation: SI, IF, and C groups: statistically significant differences

The aim of the statistical analyses that follow is to examine the existence or not of statistically significant differences as to the performance of each group of learners (SI, IF, and C) in the interpretation pre-test and in the interpretation post-test. In order to achieve this, the interpretation pre-test and the interpretation post-test data sets of the 3 groups were first submitted to the Shapiro-Wilk statistical analysis.
normality test in order to examine the null hypothesis, i.e. whether the data of each group exhibited a normal distribution pattern or not.

The results of the Shapiro-Wilk normality tests for the interpretation pre-test and the interpretation post-test are presented in Table 8:

Table 8: Interpretation: SI, IF, and C groups: Shapiro-Wilk normality tests

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group</td>
<td>Statistic</td>
</tr>
<tr>
<td>Pre-test</td>
<td>SI</td>
<td>.928</td>
</tr>
<tr>
<td></td>
<td>IF</td>
<td>.929</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>.935</td>
</tr>
<tr>
<td>Post-test</td>
<td>SI</td>
<td>.873</td>
</tr>
<tr>
<td></td>
<td>IF</td>
<td>.742</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>.941</td>
</tr>
</tbody>
</table>

The results of the Shapiro-Wilk statistical analyses for the interpretation pre-test can be described as follows: For the SI group interpretation pre-test data set the level of significance was $p=0.124$, consequently, the data in the SI interpretation pre-test data set exhibited a normal distribution pattern. For the IF group interpretation pre-test data set the level of significance was $p=0.373$, consequently, the data in the IF interpretation pre-test data set exhibited a normal distribution pattern. For the C group interpretation pre-test data set the level of significance was $p=0.563$, consequently, the data in the C interpretation pre-test data set exhibited a normal distribution pattern.
The results of the Shapiro-Wilk statistical analyses for the interpretation post-test can be described as follows: For the SI group interpretation post-test data set, the level of significance was \( p = 0.011 \), consequently, the data in the SI interpretation post-test data set did not exhibit a normal distribution pattern. For the IF group interpretation post-test data set, the level of significance was \( p = 0.002 \), consequently, the data in the IF interpretation post-test data set did not exhibit a normal distribution pattern. For the C group interpretation post-test data set, the level of significance was \( p = 0.626 \), consequently, the data in the C interpretation post-test data set exhibited a normal distribution pattern.

These data sets were further submitted to a number of analyses in order to verify hypotheses related to the existence of statistically significant differences. The tables and the analyses that follow present the results of the statistical analyses of the assessment tools used to measure the interpretation of the target structure by the 3 groups of participants / learners in the main study. 3 questions needed to be answered in 9 subsequent steps:

1. Is there a statistically significant difference between the performance of the 3 groups (i.e. SI, IF, and C) in the interpretation pre-test (steps 1 – 3)?
2. Is there a statistically significant difference between the performance of the 3 groups (i.e. SI, IF, and C) in the interpretation post-test (steps 4 – 6)?
3. Is there a statistically significant difference as to the performance of each one of the 3 groups (i.e. SI, IF, and C) between the interpretation pre-test and the interpretation post-test (steps 7 – 9)?
To answer the first question (steps 1 – 3), the following statistical analyses were performed and the pertinent results for the interpretation pre-test are presented in Table 9:

**Table 9: Interpretation pre-test t-tests comparing the SI, IF, and C groups**

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>SI-IF</th>
<th>SI-C</th>
<th>IF-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sig.</td>
<td>0.780</td>
<td>0.137</td>
<td>0.378</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.152</td>
<td>0.042</td>
<td>0.482</td>
</tr>
</tbody>
</table>

The 1st step was to compare the performance of the 21 participants in the SI instructional treatment condition in the interpretation pre-test and the performance of the 12 participants in the IF instructional treatment condition in the interpretation pre-test. Since both the SI group interpretation pre-test and the IF group interpretation pre-test data sets exhibited a normal distribution pattern, a parametric statistical analysis, i.e. the Independent Samples t-test, was performed (see Table 9).

The mean performance value of the SI group instructional treatment condition in the interpretation pre-test was 0.6 points higher than the mean performance value of the IF group instructional treatment condition in the interpretation pre-test. In the interpretation pre-test, the difference between the variance of the SI instructional treatment condition and the variance of the IF instructional treatment condition was not statistically significant (p= 0.780). The significance value of the means was p=0.152. Consequently, the difference in mean values between the SI instructional treatment condition interpretation pre-test performance and the IF instructional
treatment condition interpretation pre-test performance did not reach a level of statistical significance.

The 2nd step was to compare the performance of the 21 participants in the SI instructional treatment condition in the interpretation pre-test and the performance of the 8 participants in the C non-treatment condition in the interpretation pre-test. Since both the SI group interpretation pre-test and the C group interpretation pre-test data sets exhibited a normal distribution pattern, a parametric statistical analysis, i.e. the Independent Samples t-test, was performed (see Table 9).

The mean performance value of the SI group instructional treatment condition in the interpretation pre-test was 1.1 points higher than the mean performance value of the C group non-treatment condition in the interpretation pre-test. In the interpretation pre-test, the difference between the variance of the SI instructional treatment condition and the variance of the C non-treatment condition was not statistically significant (p=0.137). The significance value of the means was p=0.042. Consequently, the difference in mean values between the SI instructional treatment condition interpretation pre-test performance and the C non-treatment condition interpretation pre-test performance reached a level of statistical significance.

The 3rd step was to compare the performance of the 12 participants in the IF instructional treatment condition in the interpretation pre-test and the performance of the 8 participants in the C non-treatment condition in the interpretation pre-test. Since both the IF group interpretation pre-test and the C group interpretation pre-test data sets exhibited a normal distribution pattern, a parametric statistical analysis, i.e. the Independent Samples t-test, was performed (see Table 9).

The mean performance value of the IF group instructional treatment condition in the interpretation pre-test was 0.5 points higher than the mean performance value of
the C group non-treatment condition in the interpretation pre-test. In the interpretation pre-test, the difference between the variance of the IF instructional treatment condition and the variance of the C non-treatment condition was not statistically significant (p=0.378). The significance value of the means was p=0.482. Consequently, the difference in mean values between the IF instructional treatment condition interpretation pre-test performance and the C non-treatment condition interpretation pre-test performance did not reach a level of statistical significance.

To answer the second question (steps 4 – 6), the following analyses for the interpretation post-test were performed and the results of the statistical analyses executed are presented in Table 10:

Table 10: interpretation post-test test statistics comparing the SI, IF, and C groups

<table>
<thead>
<tr>
<th>Groups</th>
<th>SI - IF</th>
<th>SI - C</th>
<th>IF - C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann-Whitney U</td>
<td>119.000</td>
<td>66.500</td>
<td>42.000</td>
</tr>
<tr>
<td>Wilcoxon W</td>
<td>197.000</td>
<td>297.500</td>
<td>120.000</td>
</tr>
<tr>
<td>Z</td>
<td>-.278</td>
<td>-.892</td>
<td>-.476</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.781</td>
<td>.372</td>
<td>.634</td>
</tr>
<tr>
<td>Exact Sig. [2*(1-tailed Sig.)]</td>
<td>.811</td>
<td>.401</td>
<td>.678</td>
</tr>
</tbody>
</table>
The 4th step was to compare the performance of the 21 participants in the SI instructional treatment condition in the interpretation post-test and the performance of the 12 participants in the IF instructional treatment condition in the interpretation post-test. Since both the SI group interpretation post-test and the IF group interpretation post-test data sets did not exhibit a normal distribution pattern, a non-parametric statistical analysis, i.e. the Mann-Whitney U, was performed (see Table 10).

The mean performance value of the SI group instructional treatment condition in the interpretation post-test was 0.5 points lower than the mean performance value of the IF group instructional treatment condition in the interpretation post-test. The significance value of the means produced was p = 0.781. Consequently, the difference in mean values between the SI instructional treatment condition interpretation post-test performance and the IF instructional treatment condition interpretation post-test performance did not reach a level of statistical significance.

The 5th step was to compare the performance of the 21 participants in the SI instructional treatment condition in the interpretation post-test and the performance of the 8 participants in the C non-treatment condition in the interpretation post-test. Since the SI group interpretation post-test data set did not exhibit a normal distribution pattern, whereas the C group interpretation post-test data set exhibited a normal distribution pattern, a non-parametric statistical analysis, i.e. the Mann-Whitney U, was performed (see Table 10).

The mean performance value of the SI group instructional treatment condition in the interpretation post-test was 0.6 points lower than the mean performance value of the C group instructional treatment condition in the interpretation post-test. The
significance value of the means produced was \( p = 0.372 \). Consequently, the difference in mean values between the SI instructional treatment condition interpretation post-test performance and the C non-treatment condition interpretation post-test performance did not reach a level of statistical significance.

The 6\(^{th}\) step was to compare the performance of the 12 participants in the IF instructional treatment condition in the interpretation post-test and the performance of the 8 participants in the C non-treatment condition in the interpretation post-test. Since the IF group interpretation post-test data set did not exhibit a normal distribution pattern, whereas the C group interpretation post-test data set exhibited a normal distribution pattern, a non-parametric statistical analysis, i.e. the Mann-Whitney U, was performed (see Table 10).

The mean performance value of the IF group instructional treatment condition in the interpretation post-test was 0.1 points lower than the mean performance value of the C group instructional treatment condition in the interpretation post-test. The significance value of the means produced was \( p = 0.634 \). Consequently, the difference in mean values between the IF instructional treatment condition interpretation post-test performance and the C non-treatment condition interpretation post-test performance did not reach a level of statistical significance.

To answer the third question (steps 7 – 9), the subsequent stage was to choose and execute the appropriate statistical analyses required to accurately measure the existence or not of statistically significant differences between the interpretation pre-test and the interpretation post-test performance of each instructional treatment group (SI, IF, and C) separately. The following tables and the accompanying commentary present the results of the relevant statistical analyses.
The 7th step was to compare the performance of the 21 participants in the SI instructional treatment condition between the interpretation pre-test and the interpretation post-test. Since the SI group interpretation pre-test data set exhibited a normal distribution pattern, whereas the SI group interpretation post-test data set did not exhibit a normal distribution pattern, a non-parametric statistical analysis, i.e. the Wilcoxon Signed Ranks Test, was performed. The results are presented in Table 11 which also includes the statistical analysis, i.e. the Wilcoxon Signed Ranks Test, performed to compare the IF group interpretation pre-test performance and the IF group interpretation post-test performance:

Table 11: Interpretation SI group and IF group post-test to pre-test statistical analyses

<table>
<thead>
<tr>
<th>Post-test to pre-test test statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>SI</td>
</tr>
<tr>
<td>Z</td>
<td>-3.318</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.001</td>
</tr>
</tbody>
</table>

The mean performance value of the SI instructional treatment condition was 4.6 points in the interpretation pre-test and 3.4 points in the interpretation post-test. Since the significance value of the means produced was $p=0.001$, the difference as to the performance of the SI group between the interpretation pre-test and the interpretation post-test reached a level of statistical significance.
The 8th step was to compare the performance of the 12 participants in the IF instructional treatment condition between the interpretation pre-test and the interpretation post-test. Since the IF group interpretation post-test data set exhibited a normal distribution pattern, whereas the IF group interpretation post-test data set did not exhibit a normal distribution pattern, a non-parametric statistical analysis, i.e. the Wilcoxon Signed Ranks Test, was performed. The results have already been presented in Table 11.

The mean performance value of the IF instructional treatment condition was 4 points in both the interpretation pre-test and the interpretation post-test. Since the significance value of the means produced was p=0.787, the difference as to the performance of the IF group between the interpretation pre-test and the interpretation post-test did not reach a level of statistical significance.

The 9th step was to compare the performance of the 8 participants in the C non-treatment condition between the interpretation pre-test and the interpretation post-test. Since both the C group interpretation pre-test and interpretation post-test data sets exhibited a normal distribution pattern, a parametric statistical analysis, i.e. the Paired Samples t-test, was performed. The results are presented in Table 12:

Table 12: Interpretation C group pre-test to post-test statistical analyses

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Std. Deviation</td>
</tr>
<tr>
<td>Pre-test to Post-test</td>
<td>-6.2</td>
</tr>
</tbody>
</table>
The mean performance value of the C non-treatment condition was 3.5 points in the interpretation pre-test and 4.1 points in the interpretation post-test. Since the significance value produced was $p = 0.537$, the difference as to the performance of the C group between the interpretation pre-test and the interpretation post-test did not reach a level of statistical significance.

5.3.3. Interpretation: SI, IF, and C groups: summary

Two are the most notable findings reported in section 5.3. The first concerns the issue of distribution patterns and the second the issue of statistically significant differences. As to the former issue, in the interpretation pre-test, normal distribution patterns were documented in all 4 data sets: that of the whole sample and those of each of the SI, IF, and C groups separately. In the interpretation post-test, the only data set that exhibited a normal distribution pattern was that of the C group. As to the latter issue, the statistically significant differences documented are summarized in the following section.

5.4. Interpretation: Summary of Statistically Significant Differences

The statistically significant differences among all the data sets in the interpretation tests are summarized in Table 13:

Table 13: Interpretation: summary of statistically significant differences
<table>
<thead>
<tr>
<th>Data Set</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-test &gt; post-test</td>
<td>0.038</td>
</tr>
<tr>
<td>SI pre-test &gt; C pre-test</td>
<td>0.042</td>
</tr>
<tr>
<td>SI pre-test &gt; SI post-test</td>
<td>0.001</td>
</tr>
</tbody>
</table>

In other words, statistically significant differences were documented in 3 cases: first, between the performance of all 41 participants in the pre-test and their performance in the post-test; second, between the performance of the participants in the SI group in the pre-test and the performance of the participants in the C group in the pre-test; and third, between the performance of the participants in the SI group in the pre-test and their performance in the post-test.

5.5. Interpretation: Individual Participants’ Performance

5.5.1. Interpretation: individual participants’ performance: introduction

The following analysis presents the performance of individual participants in both the interpretation pre-test and the interpretation post-test. The aim is to examine which participants (a) improved, (b) which did not, and (c) which remained unaffected from pre-test to post-test in each of the 3 groups separately. The scale used to measure performance is in points, i.e. number 2 in the pre-test column means that this participant’s mean performance value in the pre-test was 2 points. As already mentioned in chapter 4, the performance range for both the interpretation and the production tests in the main study was from 0 to 10 points.
### 5.5.2. Interpretation: individual participants’ performance: SI group

Table 14 presents the respective performance of the participants in the SI group.

**Table 14: SI group participants’ performance in the interpretation tests**

<table>
<thead>
<tr>
<th>Group</th>
<th>Participant Number</th>
<th>Pre-test Performance</th>
<th>Post-test Performance</th>
<th>Improvement</th>
<th>Drop</th>
<th>Unaffected</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>SI</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>10</td>
<td>4</td>
<td>5</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>11</td>
<td>5</td>
<td>3</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>12</td>
<td>5</td>
<td>3</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>13</td>
<td>5</td>
<td>3</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>14</td>
<td>5</td>
<td>4</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>15</td>
<td>5</td>
<td>4</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>16</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>SI</td>
<td>17</td>
<td>6</td>
<td>2</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>18</td>
<td>6</td>
<td>3</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>19</td>
<td>6</td>
<td>4</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>20</td>
<td>6</td>
<td>4</td>
<td>√</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
According to Table 14, the performance of the 21 participants in the SI group between the interpretation pre-test and the interpretation post-test was as follows: 2 participants improved (participants number 2 and 10), 16 participants’ performance dropped (participants number 1, 3, 4, 5, 6, 7, 11, 12, 13, 14, 15, 17, 18, 19, 20, and 21), and 3 participants’ performance remained unaffected (participants number 8, 9, and 16).

5.5.3. Interpretation: individual participants’ performance: IF group

Table 15 presents the respective performance of the participants in the IF group.

Table 15: IF group participants’ performance in the interpretation tests

<table>
<thead>
<tr>
<th>Group</th>
<th>Participant Number</th>
<th>Pre-test performance</th>
<th>Post-test performance</th>
<th>Improvement</th>
<th>Drop</th>
<th>Unaffected</th>
</tr>
</thead>
<tbody>
<tr>
<td>IF</td>
<td>22</td>
<td>2</td>
<td>3</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>23</td>
<td>2</td>
<td>3</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>24</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>IF</td>
<td>25</td>
<td>3</td>
<td>7</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>IF</td>
<td>26</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>IF</td>
<td>27</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>IF</td>
<td>28</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>IF</td>
<td>29</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>IF</td>
<td>30</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td>√</td>
</tr>
</tbody>
</table>
According to Table 15, the performance of the 12 participants in the IF group between the interpretation pre-test and the interpretation post-test was as follows: 5 participants improved (participants number 22, 23, 25, 29 and 33), 6 participants’ performance dropped (participants number 26, 27, 28, 30, 31, and 31), and 1 participant’s performance remained unaffected (participant number 24).

5.5.4. Interpretation: individual participants’ performance: C group

Table 16 presents the respective performance of the participants in the C group.

Table 16: C group participants’ performance in the interpretation tests

<table>
<thead>
<tr>
<th>Group</th>
<th>Participant Number</th>
<th>Pre-test performance</th>
<th>Post-test performance</th>
<th>Improvement</th>
<th>Drop</th>
<th>Unaffected</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>34</td>
<td>1</td>
<td>6</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>35</td>
<td>2</td>
<td>2</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>36</td>
<td>3</td>
<td>2</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>37</td>
<td>3</td>
<td>5</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>38</td>
<td>3</td>
<td>7</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>39</td>
<td>5</td>
<td>3</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>40</td>
<td>5</td>
<td>4</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>41</td>
<td>6</td>
<td>4</td>
<td>√</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
According to Table 16, the performance of the 8 participants in the C group between the interpretation pre-test and the interpretation post-test was as follows: 3 participants improved (participants number 34, 37 and 38), 4 participants’ performance dropped (participants number 36, 39, 40, and 41), and 1 participant’s performance remained unaffected (participant number 35).

5.5.5. Interpretation: individual participants’ performance: summary

Finally, Table 17 summarizes the number of participants whose performance (a) improved, (b) dropped, and (c) remained unaffected between the interpretation pre-test and the interpretation post-test for each of the 3 groups, i.e. SI, IF, and C.

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of participants whose performance improved</th>
<th>Number of participants whose performance dropped</th>
<th>Number of participants whose performance remained unaffected</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI</td>
<td>2</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>IF</td>
<td>5</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

5.6. Interpretation: Item Analysis
5.6.1. Interpretation: item analysis: introduction

This section includes an analysis of how the participants in the main study responded to particular test items, i.e. tokens / verbs, in both the interpretation pre-test and the interpretation post-test. Subsection 5.6.2. presents the responses to the interpretation pre-test and subsection 5.6.3. presents the responses to the interpretation post-test. Each subsection includes an analysis of the response patterns that includes all 41 participants in the main study as well as the response patterns of the participants in each of the 3 groups, i.e. SI, IF, and C, separately. The analysis includes mean and standard deviation values. Response pattern orders are presented in separate tables followed by a short commentary in subsections 5.6.2. and 5.6.3, as well. However, before commencing the analysis, it should be noted that the distractors are not included in the analyses. This means that e.g. token 03 (distractor) is not included in both the table and the analysis that follows. Finally, mean values are reported in points and the performance range is from 0 to 10 points. For example, a mean value of 9.8 out of 10 for the token *littered* means that almost all the participants responded correctly on this particular test item. In other words, the value of 9.8 points represents the mean performance value of the participants as to the *littered* test item.

5.6.2. Interpretation: item analysis: pre-test

Table 18 presents the pertinent analysis of the interpretation pre-test.
<table>
<thead>
<tr>
<th>TOKEN</th>
<th>All groups</th>
<th>SI group</th>
<th>IF group</th>
<th>C group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>01 Littered</td>
<td>9.8</td>
<td>1.5</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>02 Carried</td>
<td>1</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>04 Played</td>
<td>1</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>05 Opened</td>
<td>4.9</td>
<td>5</td>
<td>5.7</td>
<td>5</td>
</tr>
<tr>
<td>07 Watered</td>
<td>0.7</td>
<td>2.6</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>08 Phoned</td>
<td>2</td>
<td>4</td>
<td>2.9</td>
<td>4.6</td>
</tr>
<tr>
<td>11 Coloured</td>
<td>6.8</td>
<td>4.7</td>
<td>8.1</td>
<td>4</td>
</tr>
<tr>
<td>12 Brushed</td>
<td>1.7</td>
<td>3.8</td>
<td>1.4</td>
<td>3.5</td>
</tr>
<tr>
<td>13 Deleted</td>
<td>7.8</td>
<td>4.1</td>
<td>9.5</td>
<td>2.1</td>
</tr>
<tr>
<td>15 Numbered</td>
<td>6.8</td>
<td>4.7</td>
<td>8.1</td>
<td>4</td>
</tr>
</tbody>
</table>
Beginning with all 41 participants in the interpretation pre-test, their mean performance value for particular items was in the following order: littered (mean = 9.8), deleted (mean = 7.8), coloured and numbered (mean = 6.8), opened (mean = 4.9), phoned (mean = 2), brushed (mean = 1.7), carried and played (mean = 1), and watered (mean = 0.7).

As to the SI group participants, in the interpretation pre-test, their mean performance value for particular items was in the following order: littered (mean = 10), deleted (mean = 9.5), coloured and numbered (mean = 8.1), opened (mean = 5.7), phoned (mean = 2.9), brushed (mean = 1.4), watered (mean = 1), and carried and played (mean = 0).

As to the IF group participants, in the interpretation pre-test, their mean performance value for particular items was in the following order: littered (mean = 10), deleted (mean = 7.5), numbered (mean = 5.8), coloured (mean = 5) opened (mean = 4.2), carried (mean = 2.5), played and phoned and brushed (mean = 1.7), and watered (mean = 0).

As to the C group participants, in the interpretation pre-test, their mean performance value for particular items was in the following order: littered (mean = 8.8), coloured (mean = 6.3), numbered (mean = 5), opened and deleted (mean = 3.8), played and brushed (mean = 2.5), carried and watered (mean = 1.3), and phoned (mean = 0).

The sequential order of items in the interpretation pre-test is summarized in Table 19 (mean value included in parenthesis):

Table 19: Interpretation pre-test item analysis: summary
Table 19 shows the order of tokens in various groups. The groups are ordered as follows:

<table>
<thead>
<tr>
<th>Order</th>
<th>All groups</th>
<th>SI group</th>
<th>IF group</th>
<th>C group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Littered (9.8)</td>
<td>Littered (10)</td>
<td>Littered (10)</td>
<td>Littered (8.8)</td>
</tr>
<tr>
<td>2</td>
<td>Deleted (7.8)</td>
<td>Deleted (9.5)</td>
<td>Deleted (7.5)</td>
<td>Coloured (6.3)</td>
</tr>
<tr>
<td>3</td>
<td>Coloured &amp; Numbered (6.8)</td>
<td>Coloured &amp; Numbered (6.8)</td>
<td>Numbered (5.8)</td>
<td>Numbered (5)</td>
</tr>
<tr>
<td>4</td>
<td>Opened (4.9)</td>
<td>Opened (5.7)</td>
<td>Opened (4.2)</td>
<td>Deleted (3.8)</td>
</tr>
<tr>
<td>5</td>
<td>Phoned (2)</td>
<td>Phoned (2.9)</td>
<td>Carried (2.5)</td>
<td>Played &amp; Brushed (2.5)</td>
</tr>
<tr>
<td>6</td>
<td>Brushed (1.7)</td>
<td>Brushed (1.4)</td>
<td>Played &amp;</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Carried &amp; Played (1)</td>
<td>Watered (1)</td>
<td>Phoned &amp;</td>
<td>Carried &amp;</td>
</tr>
<tr>
<td>8</td>
<td>Watered (0.7)</td>
<td>Played (0)</td>
<td>Watered (0)</td>
<td>Phoned (0)</td>
</tr>
</tbody>
</table>

According to Table 19, despite the fact that there are variations especially in the C group participants’ performance patterns, a pattern can be observed: More specifically, the tokens *littered*, *deleted*, *coloured*, and *numbered* appear in almost the same order (with the exception of the C group). The token *opened* appears always in the 5th place. As to the tokens *phoned* and *brushed*, they appear in 6th and 7th place in most cases with the exception of the C group where *phoned* appears in the 10th, final, place. The tokens *carried*, *played*, and *watered* occupy the 8th, 9th, and 10th places with between-group variations. All in all, despite variations, the sequential performance patterns of the SI and IF groups are quite similar, whereas the C group exhibits a number of distinct differences from the other 2 groups, especially as to the tokens with the lower mean values.
5.6.3. Interpretation: item analysis: post-test

Table 20 presents the pertinent analysis of the interpretation post-test.
Table 20: Interpretation post-test item analysis

<table>
<thead>
<tr>
<th>TOKEN</th>
<th>All groups</th>
<th>SI group</th>
<th>IF group</th>
<th>C group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>16 Circled</td>
<td>9.8</td>
<td>1.5</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>17 Watched</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18 Broomed</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19 Surfed</td>
<td>0.5</td>
<td>2.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21 Cooked</td>
<td>2.4</td>
<td>4.3</td>
<td>2.9</td>
<td>4.6</td>
</tr>
<tr>
<td>22 Walked</td>
<td>2.9</td>
<td>4.6</td>
<td>2.9</td>
<td>4.6</td>
</tr>
<tr>
<td>23 Danced</td>
<td>0.5</td>
<td>2.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>24 Downloaded</td>
<td>2.7</td>
<td>4.4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>27 Ticked</td>
<td>9</td>
<td>3</td>
<td>9.5</td>
<td>2.1</td>
</tr>
<tr>
<td>30 Matched</td>
<td>7.6</td>
<td>4.3</td>
<td>8.1</td>
<td>4</td>
</tr>
</tbody>
</table>
Beginning with all 41 participants in the interpretation post-test, their mean performance value for particular items was in the following order: circled (mean = 9.8), ticked (mean = 9), matched (mean = 7.6), walked (mean = 2.9), downloaded (mean = 2.7), cooked (mean = 2.4), watched and broomed (mean = 1), and surfed and danced (mean = 0.5).

As to the SI group participants, in the interpretation post-test, their mean performance value for particular items was in the following order: circled (mean = 10), ticked (mean = 9.5), matched (mean = 8.1), cooked and walked (mean = 2.9), downloaded (mean = 1), and watched and broomed and surfed and danced (mean = 0).

As to the IF group participants, in the interpretation post-test, their mean performance value for particular items was in the following order: circled and ticked (mean = 9.2), matched (mean = 7.5), downloaded (mean = 3.3), cooked and walked (mean = 2.5), watched and broomed and surfed (mean = 1.7), and danced (mean = 0.8).

As to the C group participants, in the interpretation post-test, their mean performance value for particular items was in the following order: circled (mean = 10), ticked (mean = 7.5) matched and downloaded (mean = 6.3), walked (mean = 3.8), watched and broomed (mean = 2.5), cooked and danced (mean = 1.3), and surfed (mean = 0).

The sequential order of items in the interpretation post-test is summarized in Table 21 (mean value included in parenthesis):

Table 21: Interpretation post-test item analysis: summary
<table>
<thead>
<tr>
<th>Order</th>
<th>All groups</th>
<th>SI group</th>
<th>IF group</th>
<th>C group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Circled (9.8)</td>
<td>Circled (10)</td>
<td>Circled &amp; Ticked (10)</td>
<td>Circled (10)</td>
</tr>
<tr>
<td>2</td>
<td>Ticked (9)</td>
<td>Ticked (9.5)</td>
<td>Ticked (9.2)</td>
<td>Ticked (7.5)</td>
</tr>
<tr>
<td>3</td>
<td>Matched (7.6)</td>
<td>Matched (8.1)</td>
<td>Matched (7.5)</td>
<td>Matched &amp; Downloaded (6.3)</td>
</tr>
<tr>
<td>4</td>
<td>Walked (2.9)</td>
<td>Cooked &amp; Downloaded (3.3)</td>
<td>Downloaded &amp; (3.3)</td>
<td>Downloaded &amp; (6.3)</td>
</tr>
<tr>
<td>5</td>
<td>Downloaded (2.7)</td>
<td>Walked (2.9)</td>
<td>Cooked &amp; Downloaded (2.7)</td>
<td>Walked (3.8)</td>
</tr>
<tr>
<td>6</td>
<td>Cooked (2.4)</td>
<td>Downloaded (1)</td>
<td>Walked (2.5)</td>
<td>Watched &amp; Broomed (2.5)</td>
</tr>
<tr>
<td>7</td>
<td>Watched &amp; Broomed (1)</td>
<td>Watched &amp; Broomed (1)</td>
<td>Watched &amp; Broomed (1)</td>
<td>Cooked &amp; Danced (1.3)</td>
</tr>
<tr>
<td>8</td>
<td>Surfed &amp; Danced (0.5)</td>
<td>Surfed &amp; Danced (1.7)</td>
<td>Surfed (1.7)</td>
<td>Surfed (1.7)</td>
</tr>
<tr>
<td>9</td>
<td>Surfed &amp; Danced (0.5)</td>
<td>Surfed &amp; Danced (1.7)</td>
<td>Surfed (1.7)</td>
<td>Surfed (1.7)</td>
</tr>
<tr>
<td>10</td>
<td>Surfed &amp; Danced (0.5)</td>
<td>Surfed &amp; Danced (1.7)</td>
<td>Surfed (1.7)</td>
<td>Surfed (1.7)</td>
</tr>
</tbody>
</table>

According to Table 21, a pattern can be observed: The tokens *circled*, *ticked*, and *matched* appear in the same order occupying the first 3 places for all 3 groups. The tokens *walked*, *downloaded*, and *cooked* appear in the 4th, 5th, and 6th places with between-group variations. As to the tokens *watched*, *broomed*, *surfed* and *danced*, they appear between the 7th and 10th places with between-group variations. All in all, despite variations, the performance of all 3 groups exhibits almost the same sequential performance pattern.

5.6.4. Interpretation: item analysis: summary

Two are the most notable findings of the analyses presented in section 5.6. (item analysis of the interpretation pre- and post-test). First, the participants in the 3
groups did not respond in the same way to all the items in the interpretation pre-test and to all the items in the interpretation post-test. This is evident by the significant differences as to the mean performance values for particular test items. Second, as to the interpretation pre-test, the participants in the SI group and the IF group exhibited a quite similar performance pattern, whereas the participants in the C group responded in a different way. As to the interpretation post-test, all 3 groups exhibited a quite similar performance pattern.

**Conclusion**

The results of the analyses of the interpretation tests presented in this chapter can be summarized along the following lines.

First, the distribution patterns of all 4 pre-test data sets (whole sample, SI group, IF group, and C group) were normal, whereas the distribution patterns of the post-test data sets did not exhibit normal distribution patterns, with the exception of the C group post-test data set.

Second, statistically significant differences were observed in 3 cases: (a) between the performance of all the participants in the pre-test and their respective performance in the post-test, (b) between the performance of the participants in the SI group in the pre-test and the performance of the participants in the C group in the pre-test, (c) between the performance of the participants in the SI group in the pre-test and their respective performance in the post-test.

Third, in each group individual participants responded to the in-between-test conditions, i.e. SI, IF, and C (no treatment), they were exposed to in a varied manner. More specifically, in the SI group most participants’ performance dropped, whereas in
the IF and C groups the number of participants whose performance either improved or dropped was almost the same. The number of participants whose performance remained unaffected between the pre-test and the post-test was very small for all 3 groups.

Finally, there was variation as to the way the participants responded to particular test items in both the pre-test and the post-test. In other words, the mean performance of the participants on particular test items was very varied. However, significant between-group differences as to the response patterns to the items were not documented, i.e. the performance pattern orders across groups were quite similar with minor variations. These variations were mostly documented by the C group of learners as to the interpretation pre-test items.
CHAPTER 6:

RESULTS: PRODUCTION

Introduction

This chapter includes the results of the statistical analyses as to the performance of the participants in the main study in the production pre-test and in the production post-test. The presentation of the results is divided in the following sections: In section 6.1., the mean performance values and the standard deviation values of the whole sample are included and statistically significant differences between the production pre-test data set and the production post-test data set of the whole sample are reported. In section 6.2., the mean performance values and the standard deviation values of each of the 3 groups (SI, IF, and C) separately are included and statistically significant differences as to the performance of the 3 groups separately are reported. In section 6.3., the statistically significant differences in all data sets of the production tests are summarized. In section 6.4., individual participants’ performance is analyzed. In section 6.5., an item analysis of the tokens in the pre-test and in the post-test is included. Finally, in section 6.6., an analysis of the types of errors the participants made is presented. The chapter concludes with a short summary of the most important findings of the statistical analyses as to the production of the target structure.
6.1. Production: Whole Sample

6.1.1. Production: whole sample: mean performance & standard deviation values

Table 22 presents the mean performance values and the standard deviation values of the variables production pre-test and production post-test for the whole sample of the 41 participants in the main study.

Table 22: Production pre-test and post-test mean performance values and standard deviation values (whole sample)

<table>
<thead>
<tr>
<th>MEASURE</th>
<th>PRODUCTION PRE-TEST</th>
<th>PRODUCTION POST-TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cases</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>Mean</td>
<td>2.7</td>
<td>5.5</td>
</tr>
<tr>
<td>Std</td>
<td>1.7</td>
<td>3.1</td>
</tr>
</tbody>
</table>

According to Table 22, the production pre-test mean performance value of the whole sample was 2.7 points and the production post-test mean performance value was 5.5 points exhibiting an increase of 2.7 points between the production pre-test and the production post-test. The production pre-test standard deviation value for the whole sample was 1.7 and the production post-test standard deviation value was 3.1.

6.1.2. Production: whole sample: statistically significant differences
The performance of the whole sample of 41 learners in the production pre-test and the production post-test is presented in the form of Histograms 3 and 4 respectively:

**Histogram 3:** Production pre-test performance frequencies for the SI, IF, and C groups combined (normal distribution bell curve line included)

**Histogram 4:** Production post-test performance frequencies for the SI, IF, and C groups combined (normal distribution bell curve line included)
Histogram 3 (production pre-test) clearly exhibits a clustering of performance values in the 1 and 3 values (11 and 10 cases respectively) and a negative digression in the 2 value (5 cases). There is a tendency of clustering towards the left side of the histogram, i.e. Histogram 3 presents an element of asymmetry. This could probably mean, either that the sample is not representative of the whole (possible) population of learners of this age exposed to a second language grammatical form when it comes to learning capacity, and / or that the production pre-test assessment tool used in the present study was not efficiently designed to procure a normally distributed performance by the 41 participants, i.e. that the production pre-test was quite difficult for these participants learning the target structure. A third tentative explanation might
be that the initial production of the target structure per se itself was a difficult task for these 41 learners.

Histogram 4 (production post-test) clearly exhibits small but not extensive digressions from the normal distribution line included. The 5 and 10 performance values mildly digress from the normal distribution line included but there is no clear tendency of clustering around specific values in this histogram. Histogram 4 depicts graphically that the concentration of values around the mean (5.5) is widespread, i.e. it spans quite far from the mean in the production post-test data set.

In order, however, to accurately verify or not whether the values of a data set represent a normally distributed sample, i.e. in order to verify or not the null hypothesis, graphic representations are not sufficient. They do offer initial visual information but a standard statistical analysis is also required. There are a number of available statistical analyses, but the most powerful is the Shapiro-Wilk normality test.

Consequently, the production pre-test data set and the production post-test data set (SI, IF, and C groups combined) were submitted to the Shapiro-Wilk statistical analysis normality test and the results are presented in Table 23:

**Table 23: Production pre-test and post-test normality tests (whole sample)**

<table>
<thead>
<tr>
<th></th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
</tr>
<tr>
<td>Pre-test</td>
<td>.936</td>
</tr>
<tr>
<td>Post-test</td>
<td>.930</td>
</tr>
</tbody>
</table>
The results of the Shapiro-Wilk statistical analysis can be explained as indicating a different distribution pattern between the production pre-test data set and the production post-test data set. More specifically, for the production pre-test data set the significance value was $p=0.022$. Consequently, the null hypothesis, i.e. that the production pre-test data set constituted a sample exhibiting a normal distribution pattern, was rejected. For the production post-test data set, the significance value was $p=0.15$. Consequently, the null hypothesis, i.e. that the production post-test data set constituted a sample exhibiting a normal distribution pattern, was confirmed.

The final step was to examine whether there was a statistically significant difference as to the performance of the whole sample of 41 participants (SI, IF, and C groups combined) between the production pre-test and the production post-test. Since the whole sample of the 41 participants’ production pre-test data set did not exhibit a normal distribution pattern, whereas the whole sample of the 41 participants’ post-test data set exhibited a normal distribution pattern, a non-parametric statistical analysis, i.e. the Wilcoxon Signed Ranks Test, was performed. The results are presented in Table 24:

**Table 24: Production post-test to pre-test statistical analyses (whole sample)**

<table>
<thead>
<tr>
<th>Group</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>-4.858</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>0.000</td>
</tr>
</tbody>
</table>
As aforementioned, the production pre-test mean performance value for the whole sample was 2.7 points and the production post-test mean performance value was 5.5 points. The significance value produced was $p=0.000$. In other words, the difference as to the performance of the whole sample of 41 learners who participated in the present study between the production pre-test and the production post-test reached a level of statistical significance.

6.1.3. Production: whole sample: summary

The findings in this section can be summarized along the following lines. First, the mean performance of all 41 participants increased between the pre-test and the post-test by 2.7 points and this increase reached a level of statistical significance ($p=0.000$). Second, the standard deviation value was higher in the post-test (3.1) than in the pre-test (1.7). Finally, according to the results of the normality tests, the pre-test data set did not exhibit a normal distribution pattern ($p=0.022$), whereas the post-test data set did ($p=0.15$).

6.2. Production: SI, IF, and C Groups

6.2.1. Production: SI, IF and C groups: mean performance & standard deviation values

Figure 9 presents the mean performance values as to the interpretation of the target structure of the 2 experimental groups of learners that participated in the main study (SI, IF) and the control group (C) both before the administration of the
experimental procedures (production pre-test) and after the administration of the experimental procedures (production post-test).

**Figure 9: SI, IF, and C group production pre-test and post-test mean performance values**

According to Figure 9, after being exposed to the SI experimental condition, the SI group of learners’ mean performance as to the production of the target structure increased by 2.6 points. After being exposed to the IF experimental condition, the IF group of learners’ mean performance as to the production of the target structure
increased by 3 points. The C group of learners’ mean performance as to the production of the target structure increased by 3 points.

The production pre-test and post-test data sets of each of the 3 groups of learners (SI, IF, and C) separately were further analysed and the standard deviation values produced are presented in Table 25:

Table 25: SI, IF, and C group interpretation pre-test and post-test standard deviation values

<table>
<thead>
<tr>
<th>MEASURE</th>
<th>SI Pre-test</th>
<th>SI Post-test</th>
<th>IF Pre-test</th>
<th>IF Post-test</th>
<th>C Pre-test</th>
<th>C Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std</td>
<td>1.6</td>
<td>2.8</td>
<td>1.4</td>
<td>3.7</td>
<td>1.7</td>
<td>3.3</td>
</tr>
</tbody>
</table>

According to Table 25, in the production pre-test the standard deviation value for the C group of learners was 1.7. The pre-test standard deviation value for the SI group of learners was 1.6. The pre-test standard deviation value for the IF group of learners was 1.4. In the production post-test, the standard deviation value for the IF group of learners was 3.7. The post-test standard deviation value for the C group of learners was 3.3. The post-test standard deviation value for the SI group of learners was 2.8.

6.2.2. Production: SI, IF, and C groups: statistically significant differences

The aim of the statistical analyses that follow is to examine the existence or not of statistically significant differences between the performance of each group of learners (SI, IF, and C) in the production pre-test and the production post-test separately. In order to achieve this, the production pre-test and the production post-
test data sets of the 3 groups were first submitted to the Shapiro-Wilk statistical analysis in order to examine the null hypothesis, i.e. whether the data of each group exhibited a normal distribution pattern or not.

The results of the Shapiro-Wilk normality tests for the production pre-test and the production post-test are presented in Table 26:

Table 26: Production: SI, IF, and C groups: Shapiro-Wilk normality tests

<table>
<thead>
<tr>
<th>Group</th>
<th>Shapiro-Wilk</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>df</td>
<td>Sig.</td>
</tr>
<tr>
<td>Pre-test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>.968</td>
<td>21</td>
<td>.692</td>
</tr>
<tr>
<td>IF</td>
<td>.786</td>
<td>12</td>
<td>.007</td>
</tr>
<tr>
<td>C</td>
<td>.825</td>
<td>8</td>
<td>.052</td>
</tr>
<tr>
<td>Post-test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>.933</td>
<td>21</td>
<td>.158</td>
</tr>
<tr>
<td>IF</td>
<td>.897</td>
<td>12</td>
<td>.147</td>
</tr>
<tr>
<td>C</td>
<td>.968</td>
<td>8</td>
<td>.879</td>
</tr>
</tbody>
</table>

The results of the Shapiro-Wilk statistical analyses for the production pre-test can be described as follows: For the SI group production pre-test data set the level of significance was p=0.692, consequently, the data in the SI production pre-test data set exhibited a normal distribution pattern. For the IF group production pre-test data set the level of significance was p=0.007, consequently, the data in the IF production pre-test data set did not exhibit a normal distribution pattern. For the C group production pre-test data set the level of significance was p=0.052, consequently, the data in the C production pre-test data set exhibited a normal distribution pattern.
The results of the Shapiro-Wilk statistical analyses for the production post-test can be described as follows: For the SI group production post-test data set the level of significance was $p=0.158$, consequently, the data in the SI production post-test data set exhibited a normal distribution pattern. For the IF group production post-test data set the level of significance was $p=0.147$, consequently, the data in the IF production post-test data set exhibited a normal distribution pattern. For the C group production post-test data set the level of significance was $p=0.879$, consequently, the data in the C production post-test data set exhibited a normal distribution pattern.

These data sets were further submitted to a number of analyses in order to verify hypotheses related to the existence of statistically significant differences. The tables and the analyses that follow present the results of the statistical analyses of the production assessment tools used to measure the production of the target structure by each of the 3 groups (i.e. SI, IF, and C) of participants / learners in the main study. 3 questions needed to be answered in 9 subsequent steps:

4. Is there a statistically significant difference between the performance of the 3 groups (i.e. SI, IF, and C) in the production pre-test (steps 1 – 3)?

5. Is there a statistically significant difference between the performance of the 3 groups (i.e. SI, IF, and C) in the production post-test (steps 4 – 6)?

6. Is there a statistically significant difference as to the performance of each one of the 3 groups (i.e. SI, IF, and C) between the production pre-test and the production post-test (steps 7 – 9)?

To answer the first question (steps 1 – 3), the following statistical analyses were performed and the results of the statistical analyses carried out for the production pre-test are presented in the following Tables, i.e. Table 27, Table 28, and Table 29:
Table 27: Production pre-test test statistics comparing the SI and IF instructional treatment conditions

<table>
<thead>
<tr>
<th>Groups</th>
<th>SI - IF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann-Whitney U</td>
<td>76.500</td>
</tr>
<tr>
<td>Wilcoxon W</td>
<td>154.500</td>
</tr>
<tr>
<td>Z</td>
<td>-1.895</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.058</td>
</tr>
<tr>
<td>Exact Sig. [2*(1-tailed Sig.)]</td>
<td>.063</td>
</tr>
</tbody>
</table>

Table 28: Production pre-test t-tests comparing the SI and C groups

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>SI-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sig.</td>
<td>0.921</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.048</td>
</tr>
</tbody>
</table>

Table 29: Production pre-test test statistics comparing the IF and C groups
The 1st step was to compare the performance of the 21 participants in the SI instructional treatment condition in the production pre-test and the performance of the 12 participants in the IF instructional treatment condition in the production pre-test. Since the SI group production pre-test data set exhibited a normal distribution pattern, whereas the IF group production pre-test data set did not exhibit a normal distribution pattern, a non-parametric statistical analysis, i.e. the Wilcoxon Signed Ranks Test, was performed (see Table 27).

The mean performance value of the SI group instructional treatment condition in the production pre-test was 1 point higher than the mean performance value of the IF group instructional treatment condition in the production pre-test. In the production pre-test, the difference between the variance of the SI instructional treatment condition and the variance of the IF instructional treatment condition was not
statistically significant (p= 0.058). Consequently, the difference in mean values between the SI instructional treatment condition production pre-test performance and the IF instructional treatment condition production pre-test performance did not reach a level of statistical significance.

The 2\textsuperscript{nd} step was to compare the performance of the 21 participants in the SI instructional treatment condition in the production pre-test and the performance of the 8 participants in the C non-treatment condition in the production pre-test. Since both the SI group production pre-test data set and the C group production pre-test data set exhibited a normal distribution pattern, a parametric statistical analysis, i.e. the Independent Samples t-test, was performed (see Table 28).

The mean performance value of the SI group instructional treatment condition in the production pre-test was 1.4 points higher than the mean performance value of the C group non-treatment condition in the production pre-test. In the production pre-test, the difference between the variance of the SI instructional treatment condition and the variance of the C non-treatment condition was not statistically significant (p=0.921). The significance value of the means was p=0.048. Consequently, the difference in mean values between the SI instructional treatment condition production pre-test performance and the C non-treatment condition production pre-test performance reached a level of statistical significance.

The 3\textsuperscript{rd} step was to compare the performance of the 12 participants in the IF instructional treatment condition in the production pre-test and the performance of the 8 participants in the C non-treatment condition in the production pre-test. Since the IF group production pre-test data set did not exhibit a normal distribution pattern, whereas the C group production pre-test data set exhibited a normal distribution
pattern, a non-parametric statistical analysis, i.e. the Wilcoxon Signed Ranks Test,
was performed (see Table 29).

The mean performance value of the IF group instructional treatment condition
in the production pre-test was 0.4 points higher than the mean performance value of
the C group non-treatment condition in the production pre-test. In the interpretation
pre-test, the difference between the variance of the IF instructional treatment
condition and the variance of the C non-treatment condition was not statistically
significant (p=0.039). The significance value of the means was p=0.392.
Consequently, the difference in mean values between the IF instructional treatment
condition production pre-test performance and the C non-treatment condition
production pre-test performance did not reach a level of statistical significance.

To answer the second question (steps 4 – 6), the following analyses for the
production post-test were performed and the results of the statistical analyses executed
are presented in Table 30:

Table 30: Production post-test t-tests statistics comparing the SI, IF, and C
groups

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>SI-IF</th>
<th>SI-C</th>
<th>IF-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sig.</td>
<td>0.121</td>
<td>0.626</td>
<td>0.475</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.625</td>
<td>0.414</td>
<td>0.785</td>
</tr>
</tbody>
</table>
The 4th step was to compare the performance of the 21 participants in the SI instructional treatment condition in the production post-test and the performance of the 12 participants in the IF instructional treatment condition in the production post-test. Since both the SI group production post-test data set and the IF group production post-test data set exhibited a normal distribution pattern, a parametric statistical analysis, i.e. the Independent Samples t-test, was performed (see Table 30).

The mean performance value of the SI group instructional treatment condition learners in the production post-test was 0.5 points higher than the mean performance value of the IF group instructional treatment condition learners in the production post-test. In the production post-test, the difference between the variance of the SI instructional treatment condition and the variance of the IF instructional treatment condition was not statistically significant (p=0.121). The significance value of the means produced was p= 0.625. Consequently, the difference in mean values between the SI instructional treatment condition production post-test mean performance and the IF instructional treatment condition production mean post-test performance did not reach a level of statistical significance.

The 5th step was to compare the performance of the 21 participants in the SI instructional treatment condition in the production post-test and the performance of the 8 participants in the C non-treatment condition in the production post-test. Since both the SI group production post-test data set and the C group production post-test data set exhibited a normal distribution pattern, a parametric statistical analysis, i.e. the Independent Samples t-test, was performed (see Table 30).

The mean performance value of the SI group instructional treatment condition in the production post-test was 1 point higher than the mean performance value of the C group non-treatment condition in the production post-test. In the production post-
test, the difference between the variance of the SI instructional treatment condition and the variance of the C non-treatment condition was not statistically significant (p=0.626). The significance value of the means produced was p= 0.414. Consequently, the difference in mean values between the SI instructional treatment condition production post-test mean performance and the C non-treatment condition production post-test mean performance did not reach a level of statistical significance.

The 6th step was to compare the performance of the 12 participants in the IF instructional treatment condition in the production post-test and the performance of the 8 participants in the C non-treatment condition in the production post-test. Since both the IF group production post-test data set and the C group production post-test data set exhibited a normal distribution pattern, a parametric statistical analysis, i.e. the Independent Samples t-test, was performed (see Table 30).

The mean performance value of the IF group instructional treatment condition in the production post-test was 0.4 points higher than the mean performance value of the C group instructional treatment condition in the production post-test. In the production post-test, the difference between the variance of the IF instructional treatment condition and the variance of the C non-treatment condition was not statistically significant (p=0.475). The significance value of the means produced was p=0.785. Consequently, the difference in mean values between the IF instructional treatment condition production post-test mean performance and the C non-treatment condition production mean post-test performance did not reach a level of statistical significance.

To answer the third question (steps 7 – 9), the subsequent stage was to choose and execute the appropriate statistical analyses required to accurately measure the existence or not of statistically significant differences between the production pre-test
and post-test performance of each instructional treatment group (SI, IF, and C) separately. The following tables and the accompanying commentary present the results of the relevant statistical analyses.

The 7th step was to compare the performance of the 21 participants in the SI instructional treatment condition between the production pre-test and the production post-test. Since both the SI group production pre-test data set and the SI group post-test data set exhibited a normal distribution pattern, a parametric statistical analysis, i.e. the Paired Samples t-test, was performed. The results are presented in Table 31:

### Table 31: Production SI group pre-test to post-test statistical analyses

<table>
<thead>
<tr>
<th></th>
<th>Paired Differences</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Deviation</td>
</tr>
<tr>
<td>Pre-test to Post-test</td>
<td>-2.5</td>
<td>1.8</td>
</tr>
</tbody>
</table>

The mean performance value of the SI instructional treatment condition was 3.3 points in the interpretation pre-test and 5.9 points in the interpretation post-test. Since the significance value of the means produced was p= 0.000, the difference as to the performance of the SI group between the production pre-test and the production post-test reached a level of statistical significance.

The 8th step was to compare the performance of the 12 participants in the IF instructional treatment condition between the production pre-test and the production
post-test. Since the IF group production pre-test data set did not exhibit a normal distribution pattern, whereas the IF group production post-test data set exhibited a normal distribution pattern, a non-parametric statistical analysis, i.e. the Wilcoxon Signed Ranks Test, was performed. The results are presented in Table 32:

Table 32: Production IF group post-test to pre-test statistical analyses

<table>
<thead>
<tr>
<th>Post-test to Pre-test Test Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Z</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
</tr>
</tbody>
</table>

The mean performance value of the IF instructional treatment condition was 2.3 points in the interpretation pre-test and 5.3 points in the interpretation post-test. Since the significance value of the means produced was p = 0.028, the difference as to the performance of the IF group between the production pre-test and the production post-test reached a level of statistical significance.

The 9th step was to compare the performance of the 8 participants in the C non-treatment condition between the production pre-test and the production post-test. Since both the C group production pre-test data set and the C production post-test data set exhibited a normal distribution pattern, a parametric statistical analysis, i.e. the Paired Samples t-test, was performed. The results are presented in Table 33:
Table 33: Production C group pre-test to post-test statistical analyses

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>95% Confidence Interval of the Difference</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Deviation</td>
</tr>
<tr>
<td>Pre-test to Post-test</td>
<td>-0.3</td>
<td>3.2</td>
</tr>
</tbody>
</table>

The mean performance value of the C non-treatment condition was 1.8 points in the interpretation pre-test and 4.8 points in the interpretation post-test. Since the significance value produced was $p=0.035$, the difference as to the performance of the C non-treatment condition between the production pre-test and the production post-test reached a level of statistical significance.

6.2.3. Production: SI, IF, and C groups: summary

Two are the most notable findings reported in section 6.2. The first concerns the issue of distribution patterns and the second the issue of statistically significant differences. As to the former issue, in the production pre-test, normal distribution patterns were documented in 2 data sets: those of the SI and C groups. In the production post-test, normal distribution patterns were documented in all 4 data sets: that of the whole sample and those of each of the SI, IF, and C groups separately. As to the latter issue, the statistically significant differences documented are summarized in the following section.
6.3. Production: Summary of Statistically Significant Differences

The statistically significant differences in all the data sets in the production tests are summarized in Table 34:

<table>
<thead>
<tr>
<th>Data Set</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-test &lt; post-test</td>
<td>0.000</td>
</tr>
<tr>
<td>SI pre-test &gt; C pre-test</td>
<td>0.048</td>
</tr>
<tr>
<td>SI pre-test &lt; SI post-test</td>
<td>0.000</td>
</tr>
<tr>
<td>IF pre-test &lt; IF post-test</td>
<td>0.028</td>
</tr>
<tr>
<td>C pre-test &lt; C post-test</td>
<td>0.035</td>
</tr>
</tbody>
</table>

In other words, statistically significant differences were documented in 5 cases: first, between the performance of all the participants in the pre-test and their performance in the post-test; second, between the performance of the participants in the SI group in the pre-test and the performance of the participants in the C group in the pre-test; third, as to the performance of the participants in the SI group between the pre-test and the post-test; fourth, as to the performance of the participants in the IF group between the pre-test and the post-test; and fifth, as to the performance of the participants in the C group between the pre-test and the post-test.

6.4. Production: Individual Participants’ Performance
6.4.1. Production: individual participants’ performance: introduction

The following analysis presents the performance of individual participants in both the production pre-test and the production post-test. The aim is to examine which participants (a) improved, (b) which did not, and (c) which remained unaffected from pre-test to post-test in each of the 3 groups separately. The scale used to measure performance is in points, i.e. number 4 in the pre-test column means that this participant’s mean performance value in the pre-test was 4 points. As already mentioned in chapter 4, the performance range for both the interpretation and the production tests in the main study was from 0 to 10 points.

6.4.2. Production: individual participants’ performance: SI group

Table 35 presents the respective performance of the participants in the SI group.

Table 35: SI group participants’ performance in the production tests

<table>
<thead>
<tr>
<th>Group</th>
<th>Participant Number</th>
<th>Pre-test performance</th>
<th>Post-test performance</th>
<th>Improvement</th>
<th>Drop</th>
<th>Unaffected</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI</td>
<td>1</td>
<td>4</td>
<td>9</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>√</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
According to Table 35, the performance of the 21 individual participants in the SI group between the production pre-test and the production post-test was as follows: 19 participants improved (participants number 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 19, 20, and 21), 1 participant’s performance dropped (participant number 18), and 1 participant’s performance remained unaffected (participant number 11).

6.4.3. Production: individual participants’ performance: IF group

Table 36 presents the respective performance of the participants in the IF group.
Table 36: IF group participants’ performance in the production tests

<table>
<thead>
<tr>
<th>Group</th>
<th>Participant Number</th>
<th>Pre-test performance</th>
<th>Post-test performance</th>
<th>Improvement</th>
<th>Drop</th>
<th>Unaffected</th>
</tr>
</thead>
<tbody>
<tr>
<td>IF</td>
<td>22</td>
<td>1</td>
<td>3</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>23</td>
<td>3</td>
<td>9</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>24</td>
<td>1</td>
<td>0</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>25</td>
<td>1</td>
<td>0</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>26</td>
<td>3</td>
<td>10</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>27</td>
<td>1</td>
<td>5</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>28</td>
<td>3</td>
<td>8</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>29</td>
<td>1</td>
<td>10</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>30</td>
<td>6</td>
<td>5</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>31</td>
<td>2</td>
<td>2</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>32</td>
<td>3</td>
<td>3</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>33</td>
<td>3</td>
<td>9</td>
<td>√</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to Table 36, the performance of the 12 individual participants in the IF group between the production pre-test and the interpretation post-test was as follows: 7 participants improved (participants number 22, 23, 26, 27, 28, 29, and 33), 3 participants’ performance dropped (participants number 24, 25, and 30), and 2 participants’ performance remained unaffected (participants number 31 and 32).

6.4.4. Production: individual participants’ performance: C group

Table 37 presents the respective performance of the participants in the C group.
Table 37: C group participants’ performance in the production tests

<table>
<thead>
<tr>
<th>Group</th>
<th>Participant Number</th>
<th>Pre-test performance</th>
<th>Post-test performance</th>
<th>Improvement</th>
<th>Drop</th>
<th>Unaffected</th>
</tr>
</thead>
<tbody>
<tr>
<td>C 34</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>C 35</td>
<td>4</td>
<td>10</td>
<td></td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 36</td>
<td>1</td>
<td>9</td>
<td></td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 37</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>C 38</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>C 39</td>
<td>1</td>
<td>5</td>
<td></td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 40</td>
<td>1</td>
<td>3</td>
<td></td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 41</td>
<td>1</td>
<td>6</td>
<td></td>
<td>√</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to Table 36, the performance of the 8 individual participants in the C group between the production pre-test and the production post-test was as follows: 5 participants improved (participants number 35, 36, 39, 40, and 42), 1 participant’s performance dropped (participants number 37), and 2 participants’ performance remained unaffected (participants number 34 and 38).

6.4.5. Production: individual participants’ performance: summary

Finally, Table 37 summarizes the number of participants whose performance (a) improved, (b) dropped, and (c) remained unaffected between the production pre-test and the production post-test for each of the 3 groups, i.e. SI, IF, and C.

Table 38: Summary of SI, IF and C group participants’ performance in the production tests
<table>
<thead>
<tr>
<th>Group</th>
<th>Number of participants whose performance improved</th>
<th>Number of participants whose performance dropped</th>
<th>Number of participants whose performance remained unaffected</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI</td>
<td>19</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>7</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

6.5. Production: Item Analysis

6.5.1. Production: item analysis: introduction

This section includes an analysis of how the participants in the main study responded to particular test items, i.e. tokens / verbs, in both the production pre-test and the production post-test. Subsection 6.5.2. presents the responses to the production pre-test and subsection 6.5.3. presents the responses to the production post-test. Each subsection includes an analysis of the response patterns that includes all 41 participants in the main study as well as the response patterns of the participants in each of the 3 groups, i.e. SI, IF, and C, separately. The analysis includes mean and standard deviation values. Response pattern orders are also presented in separate tables followed by a short commentary in subsections 6.5.2. and 6.5.3. Before commencing the analysis, it should be noted that there were no distractors in the production tests. Finally, mean values are reported in points and the performance range is from 0 to 10 points. For example, the mean value of 7.1 out of 10 for the token *stayed* represents the mean performance value of the participants as to this particular test item.
6.5.2. Production: item analysis: pre-test

Table 39 presents the pertinent analysis of the production pre-test.
Table 39: Production pre-test item analysis

<table>
<thead>
<tr>
<th>TOKEN</th>
<th>All groups</th>
<th>SI group</th>
<th>IF group</th>
<th>C group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Sd</td>
<td>Mean</td>
<td>Sd</td>
</tr>
<tr>
<td>1 Stayed</td>
<td>7.1</td>
<td>4.6</td>
<td>8.1</td>
<td>4</td>
</tr>
<tr>
<td>2 Cleaned</td>
<td>4.1</td>
<td>4.9</td>
<td>7.1</td>
<td>4.6</td>
</tr>
<tr>
<td>3 Looked</td>
<td>7.3</td>
<td>4.4</td>
<td>8.6</td>
<td>3.5</td>
</tr>
<tr>
<td>4 Rubbed</td>
<td>2.4</td>
<td>4.3</td>
<td>3.8</td>
<td>4.9</td>
</tr>
<tr>
<td>5 Organized</td>
<td>0.7</td>
<td>2.6</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>6 Wanted</td>
<td>0.5</td>
<td>2.1</td>
<td>0.5</td>
<td>2.1</td>
</tr>
<tr>
<td>7 Talked</td>
<td>2.2</td>
<td>4.1</td>
<td>2.4</td>
<td>4.3</td>
</tr>
<tr>
<td>8 Needed</td>
<td>0.7</td>
<td>2.6</td>
<td>0.5</td>
<td>2.1</td>
</tr>
<tr>
<td>9 Listened</td>
<td>2.2</td>
<td>4.1</td>
<td>2.4</td>
<td>4.3</td>
</tr>
<tr>
<td>10 Relaxed</td>
<td>2.4</td>
<td>4.3</td>
<td>3.3</td>
<td>4.8</td>
</tr>
</tbody>
</table>
Beginning with all 41 participants in the production pre-test, their mean performance value for particular items was in the following order: looked (mean = 7.3), stayed (mean = 7.1), cleaned (mean = 4.1), rubbed and relaxed (mean = 2.4), talked and listened (mean = 2.2), organized and needed (mean = 0.7), and wanted (mean = 0.5).

As to the SI group participants, in the production pre-test, their mean performance value for particular items was in the following order: looked (mean = 8.6), stayed (mean = 8.1), cleaned (mean = 7.1), rubbed (mean = 3.8), relaxed (mean = 3.3), talked and listened (mean = 2.4), organized (mean = 1), and wanted and needed (mean = 0.5).

As to the IF group participants, in the production pre-test, their mean performance value for particular items was in the following order: stayed (mean = 8.3), looked (mean = 5), relaxed (mean = 2.5), cleaned and talked (mean = 1.7), rubbed and organized and wanted and needed and listened (mean = 0.8).

As to the C group participants, in the production pre-test, their mean performance value for particular items was in the following order: looked (mean = 7.5), listened (mean = 3.8), stayed and talked (mean = 2.5), rubbed and needed (mean = 1.3), and cleaned and organized and wanted and relaxed (mean = 0).

The sequential order of items in the production post-test is summarized in Table 40 (mean value included in parenthesis):

Table 40: Production pre-test item analysis: summary
According to Table 40, despite the fact that there are variations in the performance patterns of the participants in all 3 groups, a pattern can be observed:

More specifically, a pattern is documented in the first 3 places and in the final 3 places. The tokens *looked* and *stayed* appear in the 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> places. The tokens *organized*, *wanted*, and *needed* appear in the final places, with the exception of the C group where the token *needed* appears in the 5<sup>th</sup> place. As to the tokens *listened*, *cleaned*, *relaxed*, *rubbed*, and *talked*, they appear in different places for each group.

Despite the sequential performance patterns of the 3 groups in the first 2 or 3 places and in the final 3 places, there are variations, especially as to the tokens with the medium mean values. Another characteristic feature of this data set is that there are

<table>
<thead>
<tr>
<th>Order</th>
<th>All groups</th>
<th>SI group</th>
<th>IF group</th>
<th>C group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Looked (7.3)</td>
<td>Looked (8.6)</td>
<td>Stayed (8.3)</td>
<td>Looked (7.5)</td>
</tr>
<tr>
<td>2</td>
<td>Stayed (7.1)</td>
<td>Stayed (8.1)</td>
<td>Looked (5)</td>
<td>Listened (3.8)</td>
</tr>
<tr>
<td>3</td>
<td>Cleaned (4.1)</td>
<td>Cleaned (7.1)</td>
<td>Relaxed (2.5)</td>
<td>Stayed &amp;</td>
</tr>
<tr>
<td>4</td>
<td>Rubbed &amp; Relaxed (2.4)</td>
<td>Rubbed (3.8)</td>
<td>Cleaned &amp;</td>
<td>Talked (2.5)</td>
</tr>
<tr>
<td>5</td>
<td>Talked &amp; Cleaned</td>
<td>Talked &amp; Organized</td>
<td>Cleaned &amp;</td>
<td>Rubbed &amp;</td>
</tr>
<tr>
<td>6</td>
<td>Listened (2.2)</td>
<td>Listened (2.4)</td>
<td>Organized &amp;</td>
<td>Needed (1.3)</td>
</tr>
<tr>
<td>7</td>
<td>Organized &amp;</td>
<td>Organized (1)</td>
<td>Wanted &amp; Organized</td>
<td>&amp;</td>
</tr>
<tr>
<td>8</td>
<td>Needed (0.7)</td>
<td>Wanted &amp; Needed</td>
<td>Wanted &amp;</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Wanted (0.5)</td>
<td>Needed (0.5)</td>
<td>Listened (0.8)</td>
<td>Relaxed (0)</td>
</tr>
</tbody>
</table>
many tokens with the same mean values, especially the ones with the lower mean values.

6.5.3. Production: item analysis: post-test

Table 41 presents the pertinent analysis of the production post-test.
Table 41: Production post-test item analysis

<table>
<thead>
<tr>
<th>TOKEN</th>
<th>All groups</th>
<th>SI group</th>
<th>IF group</th>
<th>C group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Sd</td>
<td>Mean</td>
<td>Sd</td>
</tr>
<tr>
<td>11 Loved</td>
<td>7.3</td>
<td>4.4</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>12 Helped</td>
<td>7.1</td>
<td>4.6</td>
<td>8.6</td>
<td>3.5</td>
</tr>
<tr>
<td>13 Closed</td>
<td>7.1</td>
<td>4.6</td>
<td>8.1</td>
<td>4</td>
</tr>
<tr>
<td>14 Presented</td>
<td>4.4</td>
<td>5</td>
<td>4.3</td>
<td>5</td>
</tr>
<tr>
<td>15 Raised</td>
<td>5.1</td>
<td>5</td>
<td>4.8</td>
<td>5.1</td>
</tr>
<tr>
<td>16 Liked</td>
<td>7.1</td>
<td>4.6</td>
<td>7.6</td>
<td>4.3</td>
</tr>
<tr>
<td>17 Showed</td>
<td>3.9</td>
<td>4.9</td>
<td>3.3</td>
<td>4.8</td>
</tr>
<tr>
<td>18 Clapped</td>
<td>4.9</td>
<td>5</td>
<td>4.8</td>
<td>5.1</td>
</tr>
<tr>
<td>19 Jumped</td>
<td>4.6</td>
<td>5</td>
<td>4.8</td>
<td>5.1</td>
</tr>
<tr>
<td>20 Shared</td>
<td>3.7</td>
<td>4.8</td>
<td>3.3</td>
<td>4.8</td>
</tr>
</tbody>
</table>
Beginning with all 41 participants in the production post-test, their mean performance value for particular items was in the following order: loved (mean = 7.3), helped and closed and liked (mean = 7.1), raised (mean = 5.1), clapped (mean = 4.9), jumped (mean = 4.6), presented (mean = 4.4), showed (mean = 3.9), and shared (mean = 3.7).

As to the SI group participants, in the production post-test, their mean performance value for particular items was in the following order: loved (mean = 9), helped (mean = 8.6), closed (mean = 8.1), liked (mean = 7.6), raised and clapped and jumped (mean = 4.8), presented (mean = 4.3), and showed and shared (mean = 3.3).

As to the IF group participants, in the production post-test, their mean performance value for particular items was in the following order: loved and liked (mean = 6.7), helped and raised (mean = 5.8), closed and showed and clapped and jumped (mean = 5), and presented and shared (mean = 4.2).

As to the C group participants, in the production post-test, their mean performance value for particular items was in the following order: closed (mean = 7.5), liked (mean = 6.3), helped and presented and raised and clapped (mean = 5), and loved and showed and jumped and shared (mean = 3.8).

The sequential order of items in the production post-test is summarized in Table 42 (mean value included in parenthesis):

Table 42: Production post-test item analysis: summary
<table>
<thead>
<tr>
<th>Order</th>
<th>All groups</th>
<th>SI group</th>
<th>IF group</th>
<th>C group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Loved (7.3)</td>
<td>Loved (9)</td>
<td>Loved &amp;</td>
<td>Closed (7.5)</td>
</tr>
<tr>
<td>2</td>
<td>Helped &amp;</td>
<td>Helped (8.6)</td>
<td>Liked (6.7)</td>
<td>Liked (6.3)</td>
</tr>
<tr>
<td>3</td>
<td>Closed &amp;</td>
<td>Closed (8.1)</td>
<td>Helped &amp;</td>
<td>Helped &amp;</td>
</tr>
<tr>
<td>4</td>
<td>Liked (7.1)</td>
<td>Liked (7.6)</td>
<td>Raised (5.8)</td>
<td>Presented &amp;</td>
</tr>
<tr>
<td>5</td>
<td>Raised (5.1)</td>
<td>Raised &amp;</td>
<td>Closed &amp;</td>
<td>Raised &amp;</td>
</tr>
<tr>
<td>6</td>
<td>Clapped (4.9)</td>
<td>Clapped &amp;</td>
<td>Showed &amp;</td>
<td>Clapped (5)</td>
</tr>
<tr>
<td>7</td>
<td>Jumped (4.6)</td>
<td>Jumped (4.8)</td>
<td>Clapped &amp;</td>
<td>Loved &amp;</td>
</tr>
<tr>
<td>8</td>
<td>Presented (4.4)</td>
<td>Presented (4.3)</td>
<td>Jumped (5)</td>
<td>Showed &amp;</td>
</tr>
<tr>
<td>9</td>
<td>Showed (3.9)</td>
<td>Showed &amp;</td>
<td>Presented &amp;</td>
<td>Jumped &amp;</td>
</tr>
<tr>
<td>10</td>
<td>Shared (3.7)</td>
<td>Shared (3.3)</td>
<td>Shared (4.2)</td>
<td>Shared (3.8)</td>
</tr>
</tbody>
</table>

According to Table 42, despite the fact that there are variations in the performance patterns of the participants in all 3 groups, a pattern can be observed. More specifically, a pattern is documented in the first places and in the final places. The tokens **loved, liked, helped,** and **closed** appear somewhere between the 1st and 5th places, with the exception of the C group where **loved** appears in the final place. The remaining tokens, i.e. **raised, clapped, jumped, presented, showed,** and **shared,** appear in various permutations in the final places. All in all, despite the sequential performance patterns evident, a characteristic feature of this data set is that there are many tokens with the same mean values, especially the ones with the lower mean values.

6.5.4. Production: item analysis: summary
Three are the most notable findings of the analyses presented in section 6.5. (item analysis of the production pre- and post-test). First, the participants in the 3 groups did not respond in the same way to all the items in the production pre-test and to all the items in the production post-test. This is evident in the significant differences as to the mean values for particular test items. Second, in both the pre-test and the post-test all 3 groups exhibited quite similar performance patterns with minor variations. Finally, the tokens with the lower mean values had, in many cases, the same mean values.

6.6. Production: Error Analysis

6.6.1. Production: error analysis: introduction

The present section provides a detailed error analysis of the production pre-test and the production post-test. The error patterns in the production pre-test and the production post-test are compared. As aforementioned (section 3.9.4.), scoring for the production tests was based on a 0 to 1 scoring system. 1 point was assigned to exactly correct responses (i.e. ‘loved’) or responses that were orthographically wrong but phonologically correct (i.e. ‘lovd’ or ‘lisent’). The term correct was used to refer to any answer that presented knowledge, acquisition / learning, or simple provision of the respective sound of –ed in written form as it is pronounced in the different contexts. All other erroneous responses, i.e. Present Simple Tense forms (e.g. ‘plays’), bare infinitives (e.g. ‘correct’), wrong verb use (e.g. ‘stayed’ instead of ‘loved’), and no answer, received a 0 mark. Consequently, the score range was from 0 to 10 points. There were no distractors in the production tests.
Based on this scoring procedure, the erroneous responses were further analysed. More specifically, erroneous responses were assigned to categories such as Type 1 errors (no answer / blank), Type 2 errors (wrong verb use), Type 3 errors (provision of bare infinitive form), Type 4 errors (provision of English Present Simple Tense form), and Type 5 errors (miscellaneous, i.e. including all other remaining types of errors that could not be assigned to the Type 1, 2, 3, and 4 categories, e.g. incomprehensible answers).

6.6.2. Production pre-test & post-test: error analysis: whole sample

The results of the error analyses are presented collectively for all 3 groups in the production pre-test in Table 43 and for the production post-test in Table 44.

Table 43: Production pre-test error analysis (all groups)

<table>
<thead>
<tr>
<th>All groups</th>
<th>Production pre-test error analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type 1 errors</td>
</tr>
<tr>
<td>Total number of errors</td>
<td>39</td>
</tr>
<tr>
<td>Mean</td>
<td>4.9</td>
</tr>
<tr>
<td>Std</td>
<td>2.2</td>
</tr>
<tr>
<td>Percentage of total errors</td>
<td>41</td>
</tr>
</tbody>
</table>

Table 44: Production post-test error analysis (all groups)
Production post-test error analysis (all groups)

<table>
<thead>
<tr>
<th>All groups</th>
<th>Type 1 errors</th>
<th>Type 2 errors</th>
<th>Type 3 errors</th>
<th>Type 4 errors</th>
<th>Type 5 errors</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of errors</td>
<td>27</td>
<td>9</td>
<td>15</td>
<td>7</td>
<td>19</td>
<td>77</td>
</tr>
<tr>
<td>Mean</td>
<td>3.6</td>
<td>1.1</td>
<td>1.9</td>
<td>1.4</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>2.5</td>
<td>0.3</td>
<td>1.5</td>
<td>1.1</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Percentage of total errors</td>
<td>35</td>
<td>12</td>
<td>19</td>
<td>9</td>
<td>25</td>
<td>100</td>
</tr>
</tbody>
</table>

According to Table 43 (production pre-test including all 41 participants), the error analysis exhibited the following pattern. Type 1 errors (no answer / blank) accounted for 41% of the total number of errors. Type 5 errors (miscellaneous) accounted for 26% of the total number of errors. Type 3 errors (provision of base infinitive form) accounted for 21% of the total number of errors. Type 4 errors (provision of English Present Simple Tense form) accounted for 7% of the total number of errors. Type 2 errors (wrong verb use) accounted for 5% of the total number of errors. In sum, the sequence of the types of errors beginning from the most frequent to the less frequent was as follows:

**Figure 10: Sequential pattern of types of errors in the production pre-test (all groups)**

Type 1 > Type 5 > Type 3 > Type 4 > Type 2

According to Table 44 (production post-test including all 41 participants), the error analysis exhibited the following pattern. Type 1 errors (no answer / blank) accounted for 35% of the total number of errors. Type 5 errors (miscellaneous)
accounted for 25% of the total number of errors. Type 3 errors (provision of base infinitive form) accounted for 19% of the total number of errors. Type 2 errors (wrong verb use) accounted for 12% of the total number of errors. Type 4 errors (provision of English Present Simple Tense form) accounted for 9% of the total number of errors. In sum, the sequence of the types of errors beginning from the most frequent to the less frequent was as follows:

**Figure 11: Sequential pattern of types of errors in the production post-test (all groups)**

Type 1 > Type 5 > Type 3 > Type 2 > Type 4

6.6.3. Production pre-test & post-test: error analysis: whole sample: summary

The only noteworthy difference in the error patterns between the production pre-test data set and the production post-test data sets (including all 41 participants) was that in the production pre-test Type 4 errors were in the 4th place in the sequence and Type 2 errors were in the 5th place in the sequence, whereas in the production post-test it was the other way round. More specifically in the latter case, Type 4 errors were in the 4th place in the sequence and Type 5 errors were in the 5th place in the sequence.

6.6.4. Production pre-test: error analysis: SI, IF, & C groups
The results of the error analyses are presented together for each of the 3
groups separately for the production pre-test in Table 45 and for the production post-
test in Table 46.
Table 45: Production pre-test error analysis (each group separately)

<table>
<thead>
<tr>
<th>Group</th>
<th>Type 1 errors</th>
<th>Type 2 errors</th>
<th>Type 3 errors</th>
<th>Type 4 errors</th>
<th>Type 5 errors</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI (21 participants)</td>
<td>20</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>11</td>
<td>43</td>
</tr>
<tr>
<td>Mean</td>
<td>5.1</td>
<td>1</td>
<td>1.1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>2.1</td>
<td>0</td>
<td>0.4</td>
<td>0</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>Percentage of total SI group errors</td>
<td>47</td>
<td>5</td>
<td>14</td>
<td>9</td>
<td>25 (app.)</td>
<td>100</td>
</tr>
<tr>
<td>IF (12 participants)</td>
<td>11</td>
<td>2</td>
<td>11</td>
<td>2</td>
<td>9</td>
<td>35</td>
</tr>
<tr>
<td>Mean</td>
<td>3.9</td>
<td>2</td>
<td>2.6</td>
<td>1</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>1.8</td>
<td>1.4</td>
<td>1.5</td>
<td>0</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>Percentage of total IF group errors</td>
<td>31</td>
<td>6</td>
<td>31</td>
<td>6</td>
<td>26</td>
<td>100</td>
</tr>
<tr>
<td>C (8 participants)</td>
<td>7</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>Mean</td>
<td>5.7</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>2.7</td>
<td>-</td>
<td>1.7</td>
<td>-</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Percentage of total C group errors</td>
<td>41</td>
<td>6</td>
<td>18</td>
<td>6</td>
<td>29</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 46: Production post-test error analysis (each group separately)
### Production pre-test error analysis for each group

<table>
<thead>
<tr>
<th>Group</th>
<th>Type 1 errors</th>
<th>Type 2 errors</th>
<th>Type 3 errors</th>
<th>Type 4 errors</th>
<th>Type 5 errors</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI (21 participants)</td>
<td>15</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>10</td>
<td>34</td>
</tr>
<tr>
<td>Mean</td>
<td>3.8</td>
<td>1</td>
<td>1.6</td>
<td>1</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>2.5</td>
<td>0</td>
<td>1.1</td>
<td>0</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Percentage of total SI group errors</td>
<td>44</td>
<td>12</td>
<td>9</td>
<td>6</td>
<td>29</td>
<td>100</td>
</tr>
<tr>
<td>IF (12 participants)</td>
<td>6</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>Mean</td>
<td>2.3</td>
<td>1</td>
<td>2.5</td>
<td>1.7</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>1.5</td>
<td>-</td>
<td>1.8</td>
<td>1.5</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>Percentage of total IF group errors</td>
<td>24</td>
<td>4</td>
<td>32</td>
<td>16</td>
<td>24</td>
<td>100</td>
</tr>
<tr>
<td>C (8 participants)</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>Mean</td>
<td>4.3</td>
<td>1.2</td>
<td>1</td>
<td>1</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>3</td>
<td>0.5</td>
<td>0</td>
<td>-</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Percentage of total C group errors</td>
<td>33</td>
<td>22</td>
<td>22</td>
<td>6</td>
<td>17</td>
<td>100</td>
</tr>
</tbody>
</table>
6.6.5. Production pre-test: error analysis: SI group

According to Table 45 (production pre-test), in the case of the SI group, the error analysis exhibited the following pattern. Type 1 errors (no answer / blank) accounted for 47% of the total number of errors. Type 5 errors (miscellaneous) accounted for 25% of the total number of errors. Type 3 errors (provision of base infinitive form) accounted for 14% of the total number of errors. Type 4 errors (provision of English Present Simple Tense form) accounted for 9% of the total number of errors. Type 2 errors (wrong verb use) accounted for 5% of the total number of errors. In sum, the sequence of the types of errors for the SI group in the production pre-test, beginning from the most frequent to the less frequent, was as follows:

**Figure 12: Sequential pattern of types of errors in the production pre-test (SI group)**

Type 1 > Type 5 > Type 3 > Type 4 > Type 2

6.6.6. Production pre-test: error analysis: IF group

According to Table 45 (production pre-test), in the case of the IF group, the error analysis exhibited the following pattern. Type 1 errors (no answer / blank) accounted for 31% of the total number of errors. Type 3 errors (provision of base infinitive form) also accounted for 31% of the total number of errors. Type 5 errors (miscellaneous) accounted for 26% of the total number of errors. Type 2 errors
(wrong verb use) accounted for 6% of the total number of errors. Type 4 errors (provision of English Present Simple Tense form) accounted for 6% of the total number of errors. In sum, the sequence of the types of errors for the IF group in the production pre-test, beginning from the most frequent to the less frequent, was as follows:

**Figure 13: Sequential pattern of types of errors in the production pre-test (IF group)**

Type 1 = Type 3 > Type 5 > Type 2 = Type 4

6.6.7. Production pre-test: error analysis: C group

According to Table 45 (production pre-test), in the case of the C group, the error analysis exhibited the following pattern. Type 1 errors (no answer / blank) accounted for 41% of the total number of errors. Type 5 errors (miscellaneous) accounted for 29% of the total number of errors. Type 3 errors (provision of base infinitive form) accounted for 18% of the total number of errors. Type 2 errors (wrong verb use) accounted for 6% of the total number of errors. Type 4 errors (provision of English Present Simple Tense form) also accounted for 6% of the total number of errors. In sum, the sequence of the types of errors for the C group in the production pre-test, beginning from the most frequent to the less frequent, was as follows:

**Figure 14: Sequential pattern of types of errors in the production pre-test (C group)**

Type 1 > Type 5 > Type 3 > Type 2 = Type 4
6.6.8. Production pre-test: error analysis: SI, IF, & C groups: summary

The sequential patterns of the error analysis of the SI, IF, and C groups in the production pre-test revealed that the SI and C groups exhibited an almost identical error pattern, whereas the IF group exhibited a different pattern. However, despite these differences, Type 1 errors (no answer / blank), Type 3 errors (provision of base infinitive form), and Type 5 errors (miscellaneous) accounted for the majority of errors the learners in the SI, IF, and C groups separately made in the production post-test. In contrast, Type 2 errors (wrong verb use) and Type 4 errors (provision of English Present Simple Tense form) accounted for a minority or errors the learners in the SI, IF, and C groups separately made in the production post-test.

6.6.9. Production post-test: error analysis: SI group

According to Table 46 (production post-test), in the case of the SI group, the error analysis exhibited the following pattern. Type 1 errors (no answer / blank) accounted for 49% of the total number of errors. Type 5 errors (miscellaneous) accounted for 29% of the total number of errors. Type 2 errors (wrong verb use) accounted for 12% of the total number of errors. Type 3 errors (provision of base infinitive form) accounted for 9% of the total number of errors. Type 4 errors (provision of English Present Simple Tense form) accounted for 6% of the total number of errors. In sum, the sequence of the types of errors for the SI group in the
production post-test, beginning from the most frequent to the less frequent, was as follows:

**Figure 15: Sequential pattern of types of errors in the production post-test (SI group)**

Type 1 > Type 5 > Type 2 > Type 3 > Type 4

6.6.10. Production post-test: error analysis: IF group

According to Table 46 (production post-test), in the case of the IF group, the error analysis exhibited the following pattern. Type 3 errors (provision of base infinitive form) accounted for 32% of the total number of errors. Type 1 errors (no answer / blank) accounted for 24% of the total number of errors. Type 5 errors (miscellaneous) also accounted for 24% of the total number of errors. Type 4 errors (provision of English Present Simple Tense form) accounted for 16% of the total number of errors. Type 2 errors (wrong verb use) accounted for 4% of the total number of errors. In sum, the sequence of the types of errors for the IF group in the production post-test, beginning from the most frequent to the less frequent, was as follows:

**Figure 16: Sequential pattern of types of errors in the production post-test (IF group)**

Type 3 > Type 1 = Type 5 > Type 4 > Type 2
6.6.11. Production post-test: error analysis: C group

According to Table 46 (production post-test), in the case of the C group, the error analysis exhibited the following pattern. Type 1 errors (no answer / blank) accounted for 33% of the total number of errors. Type 2 errors (wrong verb use) accounted for 22% of the total number of errors. Type 3 errors (provision of base infinitive form) also accounted for 22% of the total number of errors. Type 5 errors (miscellaneous) accounted for 17% of the total number of errors. Type 4 errors (provision of English Present Simple Tense form) accounted for 6% of the total number of errors. In sum, the sequence of the types of errors for the C group in the production post-test, beginning from the most frequent to the less frequent, was as follows:

Figure 17: Sequential pattern of types of errors in the production post-test (C group)

Type 1 > Type 2 = Type 3 > Type 5 > Type 4


The sequential patterns of the error analysis of the SI, IF, and C groups in the production post-test revealed that the SI, IF, and C groups exhibited a different pattern. These differences as to the types of errors does not permit the ascertainment of a performance error pattern in the production post-test as was the case in the production pre-test. In other words, the types of errors the learners in the SI, IF, and C
groups separately made in the production post-test were so varied that a common performance pattern cannot be verified.

6.6.13. Production: error analysis: summary

The different error patterns in all the production pre-test and the production post-test data sets are summarized in Table 47:

**Table 47: Production: error sequential patterns: summary**

**Production pre-test error sequential patterns**

<table>
<thead>
<tr>
<th>All groups</th>
<th>Type 1 &gt; Type 5 &gt; Type 3 &gt; Type 4 &gt; Type 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI group</td>
<td>Type 1 &gt; Type 5 &gt; Type 3 &gt; Type 4 &gt; Type 2</td>
</tr>
<tr>
<td>IF group</td>
<td>Type 1 = Type 3 &gt; Type 5 &gt; Type 2 = Type 4</td>
</tr>
<tr>
<td>C group</td>
<td>Type 1 &gt; Type 5 &gt; Type 3 &gt; Type 2 = Type 4</td>
</tr>
</tbody>
</table>

**Production post-test error sequential patterns**

<table>
<thead>
<tr>
<th>All groups</th>
<th>Type 1 &gt; Type 5 &gt; Type 3 &gt; Type 2 &gt; Type 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI group</td>
<td>Type 1 &gt; Type 5 &gt; Type 2 &gt; Type 3 &gt; Type 4</td>
</tr>
<tr>
<td>IF group</td>
<td>Type 3 &gt; Type 1 = Type 5 &gt; Type 4 &gt; Type 2</td>
</tr>
<tr>
<td>C group</td>
<td>Type 1 &gt; Type 2 = Type 3 &gt; Type 5 &gt; Type 4</td>
</tr>
</tbody>
</table>
Conclusion

The results of the analyses of the production tests presented in this chapter can be summarized along the following lines.

First, the distribution patterns of the both the pre- and post-test data sets were normal in the majority of cases.

Second, statistically significant differences were observed in 5 cases: (a) between the performance of all the participants in the pre-test and their performance in the post-test, (b) between the performance of the participants in the SI group in the pre-test and the performance of the participants in the C group in the pre-test, (c) as to the performance of the participants in the SI group between the pre-test and the post-test, (d) as to the performance of the participants in the IF group between the pre-test and the post-test, and (e) as to the performance of the participants in the C group between the pre-test and the post-test.

Third, the participants in each group responded to the in-between-test conditions, i.e. SI, IF, and C (no treatment), they were exposed to in a similar manner. More specifically, in all 3 groups most participants’ performance increased.

Fourth, the participants in the 3 groups did not respond in the same way to all the items in the production pre-test and to all the items in the production post-test. This is evident in the variations as to the mean performance values for particular items. However, significant between-group differences as to the response patterns to the items were not documented, i.e. the performance pattern orders across groups were quite similar with minor variations. Also, the tokens with the lower mean values had in many cases the same exact mean values.
Finally, the sequential error patterns in the production tests exhibited variations. More specifically, in the production pre-test the SI and C groups exhibited an almost identical error pattern, whereas the IF group exhibited a different pattern. In the production post-test the errors the learners in the SI, IF, and C groups made were so varied that a common performance pattern cannot be verified.
CHAPTER 7:

DISCUSSION

Introduction

The major aims of chapter 7 are (a) to assess the findings of the main study in relation to previous research both within the paradigm of studies investigating the acquisition of the English past tense as well as the paradigm of studies related with (typographical) input enhancement, input flood, processing instruction, and structured input, (b) explain the findings of the statistical analyses of our main study, (c) make tentative assumptions as to the teaching of grammar by young learners, and (d), finally, reach conclusions based on the discussion. Most of the times information from previous chapters and sections will be repeated, reiterated, or alluded to in order to provide a more coherent framework and make the discussion of the results easier. In section 7.1. the research questions are revisited. In section 7.2. the research hypotheses are revisited. In section 7.3. explanations of the results as to the interpretation of the target structure are provided. In section 7.4. explanations of the results as to the production of the target structure are provided. In section 7.5. general remarks as to both the interpretation and the production of the target structure are highlighted. In section 7.6. an attempt to provide explanations as to the performance of the control group is made. In section 7.7. the findings of the main study are
discussed within the framework of the studies related with the acquisition of the English past tense reviewed in chapter 2. In section 7.8. the findings of the main study are cross examined with a number of studies on TIE, IF, and SI reported in chapter 3. In section 7.9. implications as to teaching the regular English past tense to young Greek EFL learners validated by the findings are outlined. In section 7.10. implications as to language teaching validated by the findings are outlined. In section 7.11. implications as to the teaching of young learners are discussed. In section 7.12. implication for implicit learning validated by our findings are proposed. In section 7.13. the limitations of the present thesis are presented. In section 7.14. avenues for further research are suggested.

7.1. Research Questions Revisited

The research questions and hypotheses are repeated in order to aid cross-examination between the actual questions and the answers. Accordingly, the present thesis aimed at answering 2 main research questions:

2. Are the SI and IF & TIE implicit methodologies appropriate for the acquisition of the regular English Past Simple Tense –ed target structure?

3. Are the SI and IF & TIE implicit methodologies appropriate for the acquisition of the target structure by young Greek learners of English?

The answer to the first research question, i.e. whether implicit methodologies can aid 8-year-old Greek L1 EFL learners’ acquisition of a frequent target structure like the regular Past Simple Tense –ed, is that in all likelihood highly implicit
instructional treatments such as the simple provision of as many instances of the
target structure as possible might procure the desired results. The findings of the main
study indicate that simple exposure to the assessment measures, i.e. the control
condition, was sufficient to procure enhanced learning outcomes as to both the
interpretation and the production of the target structure. This is a valid explanation
assuming the assessment measures included simply input flooding of the target
structure. The experimental condition that combined input flooding and textual
enhancement in the form of bolding (IF), seems to have had similar beneficial effects
as to the production of the target structure but a non-effect as to the interpretation of
the target structure. The most explicit-like experimental condition, i.e. SI, seems to
have had a detrimental effect on acquisition as to the interpretation of the target
structure and a beneficial effect as to the production of the target structure. Taking
into account the enhanced learning outcomes for all groups of learners as to the
production of the target structure obviously leads to the conclusion that young
learners profit the most from highly implicit methodologies that correlate with the
abundance of input postulation in L1 acquisition. The answer then to the first research
question is that yes, implicit instruction on a frequent target structure like the regular

The second research question probed on the optimal methodology to expose
young 8-year-old Greek L1 EFL learners to a frequent target structure like the regular
English Past Simple Tense –ed. At this point, a number of tentative implications may
be provided.

The 1st tentative implication is that the age factor is not an obstacle in teaching
any target L2 structure provided that age-appropriate methodological and research
tools are designed and implemented. The findings of the main study point to the
direction that such young learners profit the most from methodologies that are highly implicit; in fact it seems that the best results are procured when these learners take part only in tests that, on their own, include input flooding. Methodologies that include elaborate instructional treatment conditions seem to hinder, or at least leave unaffected overall acquisition, especially as far as interpretation is concerned. All in all, the young learners in the present study performed the best when they were not taught, but, instead, were merely assessed, i.e. exposure to input flooded tests sufficed.

The 2nd tentative implication is that the quest for finding an optimal exposure age is not a problem if L1 and L2 acquisition are the same; it is only a matter as far as curricular decisions are concerned. This study has shown that the notion of pastness can be acquired in SLA irrespective of age. It seems that the whole issue amounts to the age when a young person can grasp the notion of pastness. All in all, the main study can simply indicate that even 8-year-old Greek L1 EFL learners can, in one way or another, understand and produce the regular English Past Simple Tense –ed. It is only a matter of designing the materials that match the age of these young learners. As a concluding remark, the age factor can be addressed successfully in teaching and learning conditions if highly implicit practices are espoused; in fact, when it comes to such young learners, it seems that the more implicit, the better.

7.2. Research Hypotheses Revisited

The research hypotheses are repeated in order to aid confirmation or not. Accordingly, the present thesis aimed at examining 2 main research hypotheses. The 2 hypotheses underlying the present study were formulated as follows:
3. Implicit instruction can help young Greek learners acquire the regular English Past Simple Tense –ed structure.

4. Young Greek L1 EFL learners learning the regular English Past Simple Tense -ed after being subjected to 2 different implicit instructional treatments will benefit the most from the instructional treatment that is the most explicit-like, and both will outperform a 3rd comparison, control group. More specifically, young Greek L1 EFL learners learning the regular English Past Simple Tense -ed after being subjected to a more explicit-like implicit instructional treatment like *structured input* (SI Group) will benefit more than young Greek L1 EFL learners learning the regular English Past Simple Tense -ed after being subjected to a more implicit instructional treatment combining *input flooding* and *textual input enhancement* (IF Group). Both groups will outperform the 3rd group (C group) that will take part only in the assessment measures without any kind of exposure to the target structure.

Research hypothesis 1 was confirmed. The results of the statistical analyses showed that implicit instruction can beneficially affect learning. This finding is relevant especially as to the production of the target structure. The statistical analyses reported significantly enhanced performance of all 3 groups in the production post-test in relation to their respective performance in the production pre-test.

As to the interpretation of the target structure the findings are not so straightforward. Although the C group of learners exhibited enhanced learning performance between the interpretation pre-test and the interpretation post-test, the same did not apply for the 2 instructional treatment conditions, i.e. SI and IF. The SI
group of learners’ performance deteriorated, whereas the IF group of learners’ performance remained unaffected. Despite an overall detrimental effect of implicit instruction as to learning in the form of the interpretation of the target structure, taking into account the performance of the C group of learners leads to the assumption that some other factor may have been operant. An attempt to explain these findings will be made in the following sections.

All in all, it is difficult to unanimously verify that implicit instruction can help young Greek learners acquire the regular English Past Simple Tense –ed structure as hypothesized by research hypothesis 1. Although the collective findings do point to this direction, the overall findings are complicated to reach definitive conclusions.

As far as research hypothesis 2 is concerned, it was hypothesized that young learners would profit the most from the most implicit type of instruction, i.e. IF, and both implicit types of instruction, i.e. both IF and SI, would consequently outperform the control group. This hypothesis was partially confirmed.

As to the interpretation of the target structure, the results of the statistical analyses showed that although the IF group of learners remained unaffected by the instructional treatment they were exposed to, the SI group of learners were detrimentally affected by the instructional treatment they were exposed to. More specifically, as initially hypothesized, the learners benefited more from being subjected to the combined input flooding and textual input enhancement instructional treatment (IF group) than being subjected to a more explicit-like instructional treatment such as structured input (SI Group). However, the issue is rather complicated due to the fact that the IF group of learners did not improve between the interpretation pre-test and the interpretation post-test, but simply remained unaffected.
As to the production of the target structure, the results of the statistical analyses showed that both the SI and IF groups of learners equally benefited from the instructional treatment they were exposed to. This means that, as to production, the hypothesis that the IF group of learners would outperform the SI group of learners was not confirmed.

On the other hand, the related hypothesis that the experimental groups would outperform the control group was not confirmed. The control group managed to perform equally as well as the other 2 experimental groups as to both the interpretation and the production of the target structure. In fact, there were no significant differences in the performance of all 3 groups of learners in both the interpretation post-test and the production post-test, despite the fact that the C group of learners performed significantly lower than the SI group of learners in both the interpretation pre-test and the production post-test. There were no significant differences in the performance of the IF group of learners and the C group of learners in any of the tests employed in the main study. In other words, the hypothesis as to the control group was not only disconfirmed, but the results suggest that it was the control group that benefited the most.

The unexpected performance of the control group will be discussed in section 7.6. However, as a concluding remark to this section, the performance of these learners further adds credence to the belief that, at this age, ample exposure, even in the form of assessment tests that include input flooding, suffices for acquisition to take place, whereas input manipulations or materials especially designed for structured input seem to distract, or even detrimentally affect learning.

7.3. Discussion of Results: Interpretation
7.3.1. Discussion of results: interpretation: introduction

This section discusses the results of the statistical analyses concerning the interpretation of the target structure as measured through the interpretation pre-test and the interpretation post-test. Section 7.3.2. includes an overall discussion of the findings Section 7.3.3. includes a discussion of the statistically significant differences. Section 7.3.4. offers some general conclusions based on the findings.

7.3.2. Discussion of results: interpretation: tentative explanations

This sub-section discusses, first, the findings of the statistical analyses as to the mean performance values and, second, as to the standard deviation values.

The discussion begins with the mean performance values. First, a possible detrimental effect as to the performance of the sub-sample of the 21 SI instructional treatment condition participants between the interpretation pre-test (mean performance value: 4.7 points) and the interpretation post-test (mean performance value: 3.4 points) can be assumed. Second, it seems that the performance of the sub-sample of the 12 IF instructional treatment condition participants between interpretation pre-test and post-test remained unaffected, since, as previously mentioned, the IF interpretation pre-test mean performance value was 4 points and the IF interpretation post-test mean performance value was 4 points as well, exhibiting an exact similarity between the interpretation pre-test and the interpretation post-test. Finally, a possible beneficial effect as to the performance of the sub-sample of the 8 C non-treatment condition participants between the interpretation pre-test (mean performance value: 3.5) and the interpretation post-test (mean performance value: 4.1)
can be assumed. In other words, the 2 experimental treatment conditions (SI, IF) and the C non-treatment condition separately examined, as regards their mean performance values, seem to have had a quite differential impact on the participants’/learners’ learning as to the interpretation of the target structure.

A number of tentative explanations for the performance of the SI group of learners can be proposed. The 1
st explanation is that SI is not an effective teaching methodology. This explanation is ruled out based on the findings of other studies reviewed in a previous chapter which have in many cases reported beneficial effects of SI. Consequently, a 2
nd related explanation might be that SI was not implemented appropriately. This explanation is also ruled out since the basic foundations of SI were adhered to as faithfully as possible. More specifically, there was no kind of explicit information in the form of a presentation of the target structure. Also, an erroneous processing strategy was focused on, i.e. the lexical preference principle, despite the fact that learners were unaware of it. Accordingly, all time markers were omitted to increase the salience of the target structure and promote noticing. Finally, the guidelines for designing SI activities were taken into account. Accordingly, at no time did the SI group of learners produce the target structure. The 3
rd explanation is that SI is not suitable for young learners as the 8-year-old pupils in the main study. This explanation seems plausible, but cannot be confirmed or disconfirmed in absolute terms unless replicated in more studies. A final explanation might be that the amount of exposure offered to the SI group was not adequate. A similar issue has been raised by Li (2012) and is presented in a subsequent section that includes possible limitations of the main study. However, it should be noted that SI instructional treatments in most studies are short and the number of activities provided are limited compared with regular classroom procedures.
To continue, tentative explanations for the performance of the IF group of learners can be proposed. The most pertinent explanation for the stagnant performance of the IF group of learners between the interpretation pre-test and the interpretation post-test is that this instructional treatment was proven ineffective. This might have been partly due to the fact that input flooding as an instructional methodology requires a more abundant amount of target tokens in order to procure the desirable results. In other words, despite the fact that the amount of tokens was accurately balanced for both instructional conditions, IF, as an instructional methodology might have required more target items in order to be effective.

As to the standard deviation values, a number of assumption may be entertained based on this measure. First, the most similar interpretation pre-test and interpretation post-test std. values were procured by the SI and C groups. In fact the pre-test and post-test std. values for the SI group almost reached the value of 1. Nonetheless, the respective values for the IF group were quite different (pre-test: 1.4, post-test: 2.3). All this can be interpreted that the experimental conditions affected each group differently. The most homogeneous response was procured by the SI group. This might have due to the large number of participants in this group. On the other hand, the relatively small number of participants in the C group renders any explanation mere speculation in the absence of an adequate number of participants in this group.

The IF group of learners exhibited a quite heterogeneous response to the condition they were exposed to. This might be attributable to 2 factors. First, the heterogeneity of the learners in this group per se. However, a most possible explanation is that input flooding as an instructional methodology, highly implicit as it is, renders possible the fact that learners respond in their own personal way to the
materials they are exposed to. In other words, the condition that included the highest element of freedom resulted in a very heterogeneous collective response.

In sum, the results of the mean values, the standard deviation values, and the normality tests might lead to a number of related tentative assumptions. First, the SI instructional treatment condition might have had a detrimental effect as for learning of the target structure affecting the whole SI group sub-sample in a quite uniform manner. Second, the IF instructional treatment condition might have had a non-effect as for learning of the target structure affecting, however, the whole IF group sub-sample in a quite non-uniform manner. Third, the C non-treatment condition might have had a beneficial effect as for learning of the target structure affecting the whole C group sub-sample in a quite uniform manner.

7.3.3. Discussion of results: interpretation: statistically significant differences

Turning now to the analyses reporting statistically significant differences as to the interpretation of the target structure, such differences were exhibited after comparing 3 data sets. First, the combined interpretation pre-test data set mean performance value was lower than the combined interpretation post-test data set (both including all 41 participants) at a level of statistical significance (p. value = 0.038). Second, the SI interpretation pre-test data set mean performance value was higher than the C interpretation pre-test data set mean performance value at a level of statistical significance (p. value = 0.042). Third, the SI interpretation pre-test data set mean performance value was lower than the SI interpretation post-test data set mean performance value at a level of statistical significance (p. value = 0.001).
All in all, the statistically significant differences of these data sets point to the tentative conclusion of a possible detrimental effect as to the interpretation of the target structure by the 41 learners who were exposed to the instructional treatment conditions.

7.3.4. Discussion of results: interpretation: general conclusions

Consequently, it seems likely that the 2 interpretation tests on their own, especially the interpretation pre-test, operated as a source of input, i.e. a form of exposure to the target grammatical structure. In all likelihood, the exposure to the interpretation pre-test and the interpretation post-test tokens of the target structure acted as a form of input flooding on its own.

All this is valid proof that the interpretation pre-test and the interpretation post-test (according to the almost uniform manner the SI and C groups of learners responded in the interpretation post-test) were indeed valid and reliable assessment tools capable of validly measuring the interpretation of the target structure by the sample of learners available in the present study.

7.4. Discussion of Results: Production

7.4.1. Discussion of results: production: introduction

This section discusses the results of the statistical analyses concerning the production of the target structure as measured through the production pre-test and the production post-test. Section 7.4.2 includes an overall discussion of the findings
Section 7.4.3 includes a discussion of the statistically significant differences. Section 7.4.4. offers some general conclusions based on the findings.

### 7.4.2. Discussion of results: production: tentative explanations

First, a possible beneficial effect in the performance of the sub-sample of the 21 SI instructional treatment condition participants between the production pre-test and the production post-test can be assumed. Second, a possible beneficial effect in the performance of the sub-sample of the 12 IF instructional treatment condition participants between the production pre-test and the production post-test can be assumed. Finally, a possible beneficial effect in the performance of the sub-sample of the 8 C non-treatment condition participants between the production pre-test and the production post-test can be assumed.

As to the standard deviation values, a number of assumptions may be entertained based on this measure. Contrary to the standard deviation values in the interpretation values, the standard deviation values in the production post-tests were much higher than the respective pre-test values. Even though the highest difference was, once again, observed in the IF group, still, very high standard deviation values were observed in the other 2 groups as well.

There are 3 possible related explanations. First, production affects individual learners in differentiated non-uniform ways unlike interpretation which procure more homogeneous responses. The second related explanation is that interpretation and production were dissociated for our learners. In other words, even though when required to interpret the target structure our learners procured more uniform group responses, when these learners were required to produce the target form, they behaved
very heterogeneously. Such an account, of course, does not discredit a third explanation, i.e. that the interpretation and the production of grammatical forms pose different levels of difficulty on individual learners.

7.4.3. Discussion of results: production: statistically significant differences

Turning now to the analyses reporting statistically significant differences as to the production of the target structure, such differences were documented after comparing the following data sets: (a) the combined production pre-test data set and the combined production post-test data set (both including all 41 participants), (b) the SI production pre-test data set and the C production pre-test data set, (c) the SI production pre-test data set and the SI production post-test data set, (d) the IF production pre-test data set and the IF production post-test data set, and (e) the C production pre-test data set and the C production post-test data set. All in all, the statistically significant differences of these data sets point to the tentative conclusion of a possible beneficial effect as to the production of the target structure by the 41 learners who were exposed to the instructional treatment conditions.

7.4.4. Discussion of results: production: general conclusions

However, the fact that the learners in the C group (non-treatment group) also exhibited the same gains in mean performance raises the issue of whether the causative variable of the beneficial effects as to the production of the target structure were due to the instructional treatments, or whether they were due to the tests themselves which might have operated as a source of additional input flood. The fact
that there were no statistically significant differences in the mean performances of the SI, IF, and C groups in the production post-test renders this possibility highly likely, especially taking into account the statistically significant higher mean performance of the SI group compared to the C group in the production pre-test.

7.5. Discussion of Results: Interpretation & Production: Summary

As to interpretation, the major finding of our main study was that there was a detrimental effect of SI on interpretation. The first explanation might be that this was due to the age of participants, i.e. learners of this age do not respond well to this instructional methodology. The second explanation was that this might have been due to an inadequacy of the research design attributed to the limited number of tokens, i.e. more activities were needed to foster acquisition (see e.g. Li, 2012, p. 425 - 427).

As to production, there are 2 noteworthy issues. First, the SI and C groups exhibited similar distribution patterns, unlike the IF group. The most likely explanation might be that IF as an instructional treatment is highly susceptible to individual variation. Second, the C group increase was almost equal to the increase of the other 2 groups. The reason behind this might have been that the enhanced performance of SI and IF groups was due to exposure to input flood in the assessment measures and not an outcome of the instructional treatments per se.

As to both the interpretation and the production of the target structure, 3 issues merit discussion. First, there was variation in the mean performance values as to interpretation for each of the 3 groups. Second, there were higher mean performance values as to production for all 3 groups. Third, there was less varied performance of learners as to interpretation than as to production.
Three related possible explanations can be proposed. First, production was easier than interpretation for these learners. Second, production is more susceptible to individual variation. Third, interpretation and production are distinct processes, at least for our learners.

7.6. Discussion: The Control Group

7.6.1. Discussion: the control group: introduction

In the main study, the performance of the C group of learners exhibited a performance pattern that merits further discussion. This applied to this group’s performance as to both the interpretation and the production of the target structure. After a brief reiteration of the major findings of the statistical analyses of the C group of learners in the present sub-section, a number of explanations will be provided in sub-section 7.6.2. A short summary of these explanations follows.

As to the interpretation of the target structure, in the interpretation pre-test the mean performance values of the respective groups were as follows: SI: 4.7 points, IF: 4 points, and C: 3.5 points. In the interpretation post-test the mean performance values of the respective groups were as follows: SI: 3.4 points, IF: 4 points, C: 4.1 points.

The difference between the C condition interpretation pre-test mean performance and the C condition interpretation post-test mean performance was not statistically significant (p. value = 0.537). The performance of the C group of learners in the interpretation post-test was higher than their respective performance in the interpretation pre-test by 0.6 points. There was a statistically significant difference between the C condition interpretation pre-test mean performance and the SI
instructional treatment condition mean performance (p. value= 0.042) (the SI group exhibited a statistically significantly higher mean performance), and there was not a statistically significant difference between the C condition interpretation pre-test mean performance and the IF instructional treatment condition interpretation pre-test mean performance. As to the interpretation post-test, there were no between-group statistically significant differences. In other words, in the interpretation post-test the C group of learners managed to surmount their statistically significant interpretation pre-test disadvantage in relation to the SI group of learners.

As to the production of the target structure, in the production pre-test the mean performance values of the respective groups were as follows: SI: 3.3 points, IF: 2.3 points, and C: 1.9 points. In the production post-test the mean performance values of the respective groups were as follows: SI: 5.9 points, IF: 5.3 points, C: 4.9 points.

The difference between the C condition production pre-test mean performance and the C condition production post-test mean performance was statistically significant (p. value = 0.035). The performance of the C group of learners in the interpretation post-test was higher than their respective performance in the interpretation pre-test by 3 points. There was a statistically significant difference between the C condition production pre-test mean performance and the SI instructional treatment condition mean performance (p. value = 0.048) (the SI group exhibited a significantly higher mean performance), and there was not a statistically significant difference between the C condition production pre-test mean performance and the IF instructional treatment interpretation mean pre-test performance. In other words, in the production post-test the C group of learners managed to surmount their statistically significant production pre-test disadvantage in relation to the SI group of learners.
7.6.2. Discussion: the control group: explanations

A number of explanations could be proposed to explain the performance of the C group of learners. The 1st explanation might be that these participants were already high achievers, i.e. that there was a measurable difference between the C group (i.e. they were already high performers) and the other 2 groups. However, as to the interpretation pre-test, there was no statistically significant difference between the C group of learners and the IF group of learners, but there was a statistically significant difference between the SI group of learners and the C group of learners (favouring the SI group). Moreover, as to the production pre-test, there was no statistically significant difference among any of the 3 groups. These facts combined rule out the possibility that the C group of learners were already more proficient than the other two groups of learners, i.e. SI and IF.

A 2nd explanation might be that the C group of learners’ performance could be attributed to cumulative test-taking effects. White commented that testing materials might have been the causative variable for the similarities in the performances of the groups in her study (White, 1998, p. 102, see also p. 103 - 104). However, this explanation is also ruled out after examination of the mean performance of the SI group of learners which declined between the interpretation pre-test and the interpretation post-test, and did not exhibit any statistically significant differences in either the production pre-test or the production post-test compared with the respective mean performances of the other 2 groups.

A 3rd tentative assumption, possibly linked with the 2nd one, might be that control conditions sometimes do exhibit similar performance patterns. For example,
similar enhanced C group performances have been reported by Marsden (2006, p. 541 – 542) and Morgan-Short & Bowden (2006, p. 57 - 58). Marsden attributed this finding to the nature of the grammar-oriented instruction incorporated in the curriculum of the school the participants were enrolled (Marsden, 2006, p. 542). Morgan-Short and Bowden attributed the C group subjects’ improved performance on (a) their exposure to the target structure, (b) test effects, (c) prior knowledge, or (d) a combination of the latter 2 points [i.e. points (b) and (c)] (Morgan-Short & Bowden, 2006, p. 57). However, it needs to be explicitly mentioned that Morgan-Short and Bowden’s control group was not a pure control group in the sense that it was provided some kind of exposure to the target form (p. 44). Despite the justified methodological objections this exposure raises as to the internal validity of their whole research design, it is a fact that Morgan-Short and Bowden’s control group and the control group in the our main study differed fundamentally since the latter received no exposure at all as to the target structure. Consequently, their findings and ensuing argumentation and explanation can not be applied in the present case.

Farley and Aslan also reported enhanced learning outcomes by a control group in their study within the PI framework. In their study, as to interpretation, despite enhanced learning outcomes for both experimental groups, i.e. PI and MOI, the enhanced learning outcomes of the control condition implied that some other factor, other than the treatment conditions, may have been operable. In Farley and Aslan’s own words, since “the Control group improved, it is hard to claim that the improvement made by the PI and MOI groups is due to their relative treatments” (Farley & Aslan, 2012, p. 131). On the other hand, as to the production of the target structure, “the Control group did not make any improvement”, whereas both the PI and MOI groups exhibited enhanced learning outcomes (p. 133). This is diametrically
different from the main study in the present thesis where the control group exhibited enhanced learning outcomes as to both interpretation and production

The performance of the control condition “remains a mystery” for Farley and Aslan (p. 135). The 2 authors attributed it to the proficiency level or test-taking effects, which both seem equally valid. However, they also attributed this fact to learners’ prior knowledge but insisted that “the improvement could not be attributed to any prior knowledge that the participants in the Control group had since if that had been the case, those participants would have showed, at minimum, some improvement on the production tasks” (p. 135). The objection is that these participants may have not been provided the appropriate tools to improved production. In other words, the issue might not have been due to a learner trait, i.e. low prior knowledge, but not sensitive enough production tools, i.e. it might have been due to a methodological feature incorporated in their research design.

A 4th relevant tentative explanation might be that the assessment measurers incorporated in the main study of the present thesis operated as a form of input enhancement on their own. In fact, a close inspection of the interpretation and the production tests reveals (a) that the interpretation tests (pre-test and post-test combined) included exposure to 20 target form tokens (plus 10 distractors), and (b) that the production tests (pre-test and post-test combined) included exposure to 20 target form tokens, too. In other words, it could be tentatively assumed that the assessment measures included, on their own, input flooding to the target structure. Taking into account the fact that the total number of available tokens was relatively small, i.e. 55, (due to the age and the proficiency level of the young participants), easily leads to the conclusion that the C group of learners received a relatively high amount of input flooding that included the majority of available tokens.
It can be, therefore, tentatively assumed that input flooding and textual input enhancement (in the form of bolding) for the IF treatment condition might have had a detrimental effect as to the interpretation of the target structure. On the other hand, the C condition was only exposed to input flooding in the interpretation test and the production tests. The aforementioned combination for the IF group might have caused a trade-off effect on comprehension for the IF treatment condition (see Han, Park & Combs, 2008, p. 603 for details and studies where a trade-off effect of textual enhancement on comprehension might have been operant).

Consequently, the combination of input flooding and textual enhancement might have been the causative variable for the findings in the IF group, especially for the young learners in the main study. In other words, textual enhancement might have negatively affected the hypothesized beneficial effects of input flooding as Han et al. also suggested (p. 610). Textual enhancement has been, in fact, evaluated as being more explicit than input flooding (White, 1998, p. 86). Assuming that the C group of learners only received input flooding in the form of the assessment tokens, the previous argumentation (4th explanation) seems a valid assumption.

This point has been discarded in both the theoretical framework and the current research carried out in the field of SLA. It is especially pronounced in the studies carried out within the PI framework where the majority of studies are based on interpretation tests and production tests very similar to the ones employed the main study. The fact that very few studies have been carried out with such young learners and the consequent availability of more tokens in these studies does not cancel out this line of reasoning, since exposing older and more proficient learners, even adolescents, to input flooding, coupled with whatever type of instruction, might be a
tentative causative variable for the reported enhanced learning outcomes of treatment conditions.

As a matter of fact, the way the control non-treatment condition, whenever used, has been implemented is not explicitly mentioned and the majority of studies that do use one simply report a control treatment without providing details even on basic statistical issues such as the presence or absence of normal distribution patterns in the data. To be more precise, normal or abnormal distribution patterns are hardly ever reported even though they are a prerequisite for further statistical analyses (e.g. Norris & Ortega, 2006, qtd. in Spada & Tomita, 2010, p. 276).

A 5th relevant set of explanations might be linked with working memory. First, (a) structured input might have overloaded the short-term memory of the SI group of learners, (b) input flooding and textual enhancement might have presented a medium-level load for the short-term memory of the IF group of learners, and (c) the unfocused input flooding of the assessment measures might have limited the working memory load for the C group of learners. This explanation is highly likely for the young learners in the main study whose long-term memory of the target language was largely empty, i.e. they lacked L2 long-term memory necessary to retrieve knowledge from such a system in order to make relevant L1 - L2 comparisons, something which “requires selective attention as well as the executive control of what is activated so that limited attentional resources will not be overwhelmed” (Mackey, Adams, Stafford, & Winke, 2010, p. 519). This overload explanation might be valid as to the interpretation of the target structure (see also Leow, Hsieh, & Moreno, 2008, p. 687 for the effects of attentional overload).

However, the concept of working memory might also explain the enhanced learning outcomes of all 3 groups as to the production of the target structure.
Accordingly, these 41 young learners might have photographically reproduced the target items in the production tests merely copy-pasting the tokens. Robinson refers to this copy-pasting as infants’ “copying capacity” (Klein & Dimroth, 2009, qtd. in Robinson, 2010, p. 260). This might have also been aided by the use of relevant photographs which accompanied the materials in the present study. Young children are especially capable of orally reproducing language chunks without consciously understanding precisely their meaning. There is no reason to refute a similar phenomenon in written language if such young L2 learners have mastered hand coordination, and production of letters and words or language chunks. In fact, it has been claimed that learners often parrot, and / or repeat feedback (Gass, 2006, qtd. in Mackey et al., 2010, p. 520). Although the design of the main study did not include feedback as a variable, there is reason to believe that such parroting might have characterised learners’ production of the target form, especially since they only had to reproduce target words they had both seen and heard. In fact, as Robinson suggested, even though young learners lack “entrenched pre-existing knowledge of L1 or other language constructional form-meaning mappings”, they are capable of noticing frequent “co-occurring stimuli in the input … (that support) … generalization and learning of structure in the absence of developed WM and long-term memory abilities” (Elman, 1993; Goldberg, 2006; Newport, 1990, qtd. in Robinson, 2010, p. 260, parenthesis added).

This explanation is further substantiated taking into account the fact that the materials the learners were exposed to in both treatment conditions, and the assessment measures (the same for all 3 conditions), were always based on familiar topics and accompanied by age-appropriate photographs. Despite their young age then, these learners did have the sufficient background knowledge required to activate
topic-related *scripts* (Schank & Abelson, 1977; Cullingford, 1978, qtd. in Saeed, 1997, p. 186) that facilitate learning/acquisition. In other words, they were in a position to activate relevant cognitive schemata through topic familiarity, a factor that has been reported as beneficial for the establishment of form-meaning connections (Leeser, 2003, p. 254) by lessening working memory load (p. 230-234).

If such an interpretation based on working memory is valid, then, in all likelihood, comprehension and production are distinct processes. This is certainly borne out by the findings of the main study in which the performance of the 41 young learners on the whole deteriorated as to the interpretation of the target structure, but increased as to the production of the target structure. The fact that a *parroting* effect might be the causative variable for production does not invalidate this line of thought since mimicry is an essential characteristic of child learning in general. In this case, 2 options seem feasible. Either the assessment tests were not accurate enough to measure comprehension and understanding, or that production is easier than comprehension. This final explanation is contrary to the standard assumption of the Herbartian-oriented PPP model where production is last. Maybe it is easier to (re)produce language than comprehend it, especially for young learners like the ones in the main study.

A 6th relevant point worthy of detailed discussion concerns the cut-off scores and the statistical analyses utilized in most research designs within the broader framework of explicit and implicit instruction. The methodological reason behind arbitrarily using cut-off scores is the attempt to ensure that subjects have little prior knowledge of a target structure. Radwan, who noted that in his study “no cut off score was set for the elimination of any subjects” (Radwan, 2009, p. 281), similar to the main study, is a notable exception. This is a marked contrast in relation to other IP-
related studies where a 60% or 70% cut-off score has been commonly applied. Radwan’s decision was motivated by the fact that form-focused instruction might procure detrimental learning effects (Felix, 1981, Lightbown, 1983, qtd. in Radwan, 2009, p. 281). This is a crucial issue and Radwan’s methodological motivation is valid since it renders possible the fact that very high pre-test performers, e.g. in the 80% to 90% range, may be affected negatively by an instructional treatment condition. As a matter of fact, a 60% cut-off score ensures that only medium proficiency learners are included in the participant pool, and, consequently, there is at least a 40% statistical likelihood that learners will benefit from instruction. On the other hand, the elimination of very high performers minimizes the chance that these learners will be unaffected or detrimentally affected by instruction. After all, their chances of improvement are minimal.

At the same time, such cut-off scores create very homogeneous language learner groups not representative of real life conditions, where statistically within a span of approximately ± 3 standard deviations the whole population is to be included. In other words, cut-off scores create non-representative sample populations and, in a way, impose stabilized or enhanced learning outcomes. For example, a cut-off score of 60% is about ± 2 standard deviations which implies that any generalization of the results to the whole population cannot be substantiated since another 40% is excluded from the statistical analyses.

Such arbitrary cut-off scores cannot be sustained methodologically, especially if one takes into account an old formula of elementary statistics, i.e. the Poisson Law. According to Poisson, the higher the performance, the narrower the scope of improvement, and vice versa; if performance is low / small, there is a higher chance of improvement. Based on random trialling (similar to throwing the dice), the Poisson
Law assumingly predicts the following: if an imaginary subject scores 80% in the pre-test, there is a 19% chance of scoring higher and a 79% chance of scoring lower in the post-test.

The sceptics concerned about random trialling should be reminded that the statistical analyses and results of our studies are possibilities based on arbitrary cut-off points, which are in turn based on assumptions concerning probabilities / possibilities. In other words, excluding learners from a participant pool on the basis of their very high or very low performance is similar to tampering with the dice so as to procure the desired results. As Leow et al. reported, if they had not used a 60% cut-off score in their study, their findings would have been different (Leow et al., 2008, p. 685).

That is to say, a subject scoring e.g. 70% on a pre-test might score less in the post-test, e.g. 50%. By excluding this imaginary subject, the fact that a given instructional treatment may be detrimental for learning for this particular subject is ignored as there is an e.g. 69% chance for such a state of affairs within the normal distribution area of the total subject pool as the following diagram displays:

Diagram 1: Likelihood of post-test performance of imaginary subject who scored 70% in the pre-test
All things then considered, it is safe to argue that the use of young participants is indeed ideal, at least for studying the English regular Past Simple Tense -ed since minimal initial knowledge is not only expected but desirable as well.

Merging Poisson’s random trialling and Gaussian Bell curves can lead to a re-examination of standardized analyses. For example, juxtaposing the mean of e.g. Group A, B, or C based on the combined pre-test and post-test results might produce a pictorial pattern offering differentiated interpretations. In the case of the main study, a re-interpretation of the data could be proposed. More specifically, instead of using the mean pre-test performance and the mean post-test performance of the SI, IF, and C groups respectively as reference points, we could tentatively use the combined mean pre-test and post-test performance of each of the 3 groups and set as a reference point the combined mean pre-test and post-test performance of the Control group. In this situation, the SI Group would be designated a line on the diagram, the IF Group still another line, and the C Group yet another line acting as the reference point. Such is
the state of affairs, if the fact that the SPSS analyses produced quite normal
distribution pre-test and post-test distribution patterns is taken into account. In fact,
the importance of including histograms and distribution patterns has not been
adequately addressed in the relevant literature so far. “Sample sizes in applied
linguistics are usually small and the homogeneity test is likely to turn out to be
significant” and this necessitates graphic distribution checks (Norris & Ortega, 2006,
qtd. in Spada & Tomita, 2010, p. 276). The result could appear something like the
following diagram:

Diagram 2: Imaginary normal distribution diagram including all subjects and
using combined pre-test and post-test mean scores as reference points
The above imaginary diagram (merging Poisson’s random trialling and Gaussian Bell curves) could pictorially present the true scope of improvement of the SI and IF Groups. All the subjects are to be included irrespective of their pre-test scores. The inclusion of subjects which might have been eliminated (very high pre-test achievers, i.e. high outliers) could assumedly very tentatively explain the results of the decreased SI Group performance and the stagnant IF Group performance in the post-tests compared to their respective performance in the pre-tests.

Finally, a 7th tentative explanation might be linked to affective variables. More specifically, the present researcher was the instructor in the SI and IF treatment conditions and the assessment measures. The participants did not know him at all. Also, these young learners knew they were taking part in an experiment and assessed by a researcher since they were probably familiar with this information due to the fact that parental consent was explicitly required and granted in writing. However, whereas the SI group of learners and the IF group of learners completed the treatment materials under the researcher’s guidance, the C group of learners were occupied with study-irrelevant activities by one of the regular English teachers of the school. This fact might have contributed to increasing stress and anxiety for the SI and IF groups of learners, whereas the C group might have felt more comfortable and at ease. All in all, the affective climate between, on the one hand, the treatment conditions, and, on the other hand, the control condition, might have been diametrically different.

7.6.3. Discussion: the control group: summary

The explanations proposed in section 7.6.2. to explain the performance of the control group in the main study are summarized in Table 48:
### Table 48: Summary of explanations of C group performance

<table>
<thead>
<tr>
<th>No</th>
<th>Explanation</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>The C group participants were already high achievers.</td>
<td>Ruled out by statistical analyses.</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>Test-taking effects.</td>
<td>Ruled out based on the performance of the other 2 groups.</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>Control groups sometimes exhibit similar performance patterns.</td>
<td>Ruled out since control groups that have exhibited similar performance patterns differed from main study control group as to their implementation and incorporation in the respective research designs.</td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt;</td>
<td>Assessment measures operated as a form of input enhancement including input flooding.</td>
<td>Ruled out since it should have beneficially affected the performance of the other 2 groups. Sustained since input flooding was unintended but present in the design of the assessment measures.</td>
</tr>
<tr>
<td>5&lt;sup&gt;th&lt;/sup&gt;</td>
<td>Working memory.</td>
<td>Sustained since (a) memory load was not the same for all groups, (b) a <em>parroting effect</em> as to production might have been operant. It might be tentatively assumed that interpretation and production are distinct processes.</td>
</tr>
<tr>
<td>6&lt;sup&gt;th&lt;/sup&gt;</td>
<td>Cut-off scores.</td>
<td>Sustained and elaborated on.</td>
</tr>
<tr>
<td>7&lt;sup&gt;th&lt;/sup&gt;</td>
<td>Affective variables.</td>
<td>Sustained.</td>
</tr>
</tbody>
</table>

### 7.7. Discussion: The Acquisition of the English Past Tense

#### 7.7.1. The acquisition of the English past tense: introduction

In this section we will attempt to relate our findings and research agenda in general with studies that have investigated the acquisition of English past tense by either English L1, or English L2 learners. Only a small number of observations will
be reported since these studies did not include similar research designs as ours and their goals are very different than ours. In fact, the research design of our main study is markedly different from the research designs of these studies. In this respect, it is invalid to cross-examine our study with these studies and reach valid conclusions on this basis. For this reason, only a small number of pertinent issues raised by such studies in section 2.5. will be alluded to in order to contribute to a better understanding of the findings in our study.

7.7.2. Discussion: The acquisition of the English past tense: pertinent remarks

The fundamental difference between our study and these studies is that the vast majority of these studies cross-examined the acquisition of both the regular and the irregular forms of the English past tense, whereas we studied only the regular English past tense form. This renders any generalizations invalid. However, there exist a number of points that could aid our discussion.

To begin with, Kuczaj reported that regular past tense forms are acquired earlier than irregular forms and that overgeneralization errors vary with age. Despite the fact that Kuczaj makes no link of this finding to regular past tense forms, there might be a connection with such forms. If children begin by over-generalizing –ed to base forms, this might signify that the correct or incorrect addition of this suffix to regular base forms might depend either on properties of the base forms themselves, or abundant or non-abundant presence of such forms in the input. In the first case, some regular past tense verb forms might be more easily acquired due to inherent phonological, morphological, and syntactical attributes of the base verb forms. In the second case, it seems logical to argue that as children grow older their exposure to
verb forms increases leading to higher levels of regular past tense acquisition. All younger children have to rely on is the frequency of past tense forms in the input they receive. This has direct implications for foreign language teaching as to the selection of regular verb forms in the input. In other words, (young) learners will (a) more easily acquire some past tense forms rather than others based on the inherent properties of the verb forms themselves, and (b) they will acquire more easily verbs forms present in the input provided. It seems then that the overgeneralization patterns of irregular past tense forms may partly explain regular past tense acquisition patterns based on Kuczaj’s findings.

Bybee and Slobin (1982) compared the schemas used by their 3 age groups and this comparison was also applied to English regular past tense forms. This can be of relevance to the present main study. More specifically, Bybee and Slobin argued that (a) preschoolers apply both source-oriented and product-oriented modifications when forming regular past tense forms and (b) adults and 3rd-graders also apply both types of modifications, albeit in a more fine-tuned manner (p. 286 - 287). According to the 2 authors, if preschoolers erroneously classify a regular past tense form ending in /t/ or /d/ to the product class, they may not supply the –ed suffix. Such an account implies that when these children supply base regular forms without the appropriate –ed suffixation, this error does not necessarily constitute absence of learning, but, instead, may be interpreted as a misapplication of the appropriate schema due to morphological properties of the past tense form itself. Non-provision of the appropriate suffixation for some regular past tense forms may not constitute an error per se, but, rather, a developmental marker of the learners’ current interlanguage (IL) state. Bybee and Slobin argued that adults and 3rd-graders are not as susceptible as preschoolers to the same erroneous strategies. The significance of such observations
in relation to the main study reported in the present thesis is that the provision and consequent classification of regular base forms instead of the correct –ed suffixed forms may not necessarily constitute an error but rather a developmental stage attribute. In other words, not all base form provision errors are the same type of error, but a more fine-tuned classification system based on morphological properties of the base form may be required.

The notions of type frequency and token frequency raised by Bybee (1995) are of particular relevance for our study. In all likelihood, the performance of the young children in the main study is not solely attributable to the (instructional) conditions they were exposed to. The type and token frequency of the tokens the participants were required to interpret and produce might have influenced performance on particular test items thereby affecting the overall efficacy of the teaching methodologies assessed. This necessitates that a detailed item analysis that takes into account type and token frequency might explain why participants performed differentially on particular test items. Test tokens of high type frequency relations and high token frequency should, according to Bybee’s network model account, be more easily acquired than low type frequency relations among test items and low token frequency test items. In other words, the performance of the young children in the main study might not be a result of the (instructional) conditions per se, but might have been influenced by inherent phonological, semantic, and morphological connections among target test items, as well as by the frequency of such items in the input they had already been exposed to after 3 years of EFL study.

The findings of Rice et al. (1995) and Rice and Wexler (1996) can also corroborate with our present purposes. It is possible that the classification of bare stem forms, instead of the appropriate –ed suffixed form, as errors may not be as
straightforward and accurate as expected. For example, it may be necessary that we examine such errors for individual subjects separately. In this way, individual subjects who rely more than other subjects on this uniflected strategy may be discovered. It is unwarranted then both by our research agenda and by our research purposes to label such subjects as linguistically impaired. In our case, a more accurate explanation of such findings may be interpreted that a number of subjects in our participant pool may belong to an earlier developmental stage than their peers. Such an interpretation is also inferable by Rice et al.’s discussion as to how impairment may be manifested (Rice et al., 1995, p. 862).

The findings of Jaeger et al. also have relevance to the present main study. It seems that regular past tense forms are easier to process rendering them more appropriate than irregular past tense forms for young learners since the former procure smaller RTs and require less brain activation. If this line of reasoning is correct, erroneous performance on regular past tense forms might offer more valuable insights than correct performance on irregular past tense forms. This is due to the fact that correct performance per se can be interpreted simply as acquisition. On the other hand, erroneous performance can be interpreted as (a) absence of acquisition, (b) inability to respond at an expected pre-specified amount of time, or (c) even deficiencies in the brain. Farfetched though it may seem, according to this account, the processing and acquisition of regular past tense forms is not simply a matter of learning; it has indirect psychological and neurological implications as Jaeger et al. confer in their discussion of their findings.

A more detailed analysis of the error patterns reported by Marchman might be relevant to the findings of the present main study. Errors were reported as to all past tense verb forms with the exception of try (p. 292). The most frequent regular past
tense form errors were incorrect zero-markings and vowel changes (p. 291). As to zero-marking errors, these were more frequent when the base stem form ended in alveolar consonants. In fact, Marchman found out that all verbs, both regular and irregular, that never exhibited zero-marking errors did not end in the consonants /t/ or /d/ (p. 297). The presence of a dominant vowel in the base stem form also resulted in zero-markings (p. 298). On the basis of such evidence, it can be argued that zero-marking errors reported in the present main study may not be attributable to inefficacies of the (treatment) conditions our young children were exposed, but, rather to inherent morphophonological properties of the target test items themselves.

Marchman also drew a conclusion of high, albeit indirect, relevance to our main study. She stated that the productivity patterns of regular and irregular past tense forms share more similarities than differences. In her own words,

“… past tense findings are consistent with the conclusion that the mechanisms underlying the productive use of regular and irregular patterns are actually more similar than dissimilar.” (Marchman, 1997, p. 300)

This statement can be interpreted as implying that the error patterns and the explanations to account for such patterns in our study incorporating exclusively regular past tense verbs might be generalized to the acquisition of irregular past tense verbs. The fact that Mrachman limited her conclusion to productivity patterns still renders such an interpretation valid. However, most theoretical, research, and in-practice accounts are based on a possible dissociation between these 2 types of forms.

The issues of error patterns and imageability reported by Miozzo (2002) may offer some tentative implications for our study. The first issue is related with the fact that Miozzo’s (2002) patient produced very few errors with regular past tense forms in the following order from more to less frequent: (a) provision of base stem form, (b) addition of the suffix –s to the base stem form, and (c) change of stem. In our study,
we found evidence of all of these 3 types of errors. Miozzo attributed this very limited number of errors to the patients’ unimpaired accessibility to lexical semantics, i.e. to the patient’s “intact access to word meaning” (p.116). This can be taken as evidence suggestive that our young participants were not linguistically impaired or, to be more accurate, had no problem accessing the semantics of the tokens, i.e. English regular past tense verbs. In fact, the provision of the base stem form seems to operate as what Miozzo labeled a “default” strategy (p. 117). It is possible that the young children in our study also relied on such a strategy in many cases. It is difficult to argue convincingly that the provision of such base stem forms is an indication of absence of past tense acquisition or a simple inability to supply the appropriate –ed suffix in surface form as Rice et al.’s (1995), and Rice and Wexler’s (1996) (Extended) Optional Infinitive Stages suggest.

As to the latter issue, i.e. that of imageability, Miozzo defined it as “the ease with which people can form a sensorial image of the concept expressed by a word” (Miozzo, 2002, p. 118). Miozzo addressed imageability only as to English irregular past tense forms dividing them in 2 categories: high-imageability verbs and low-imageability verbs. The author reported that imageability had a relatively small effect on the patient’s overall performance, i.e. the patient performed slightly better with verbs of high imageability (p. 118 – 119). Issues related with imageability were of concern in our study, as well. More specifically, we faced the problem of shooting photos for a limited number of verbs due to the fact that they were not easy to depict in pictorial form, e.g. tokens such as needed, wanted, loved, etc. We suspected that inability to accurately capture in pictorial form the meaning of such verbs might have negatively affected subjects’ performance on these verbs. However, Miozzo’s findings and discussion of imageability indicate that what is important is not
imageability per se, but access to the meaning of the verbs. It should be underscored that our subjects both saw pictures, and read and heard sentences for all of the tokens included in the assessment tasks. In other words, low performance on such difficult to depict verbs is more attributable to the fact that our young learners did not know the meaning of such tokens, rather than to our inability to accurately depict the meaning of these tokens in pictorial form. Despite differences, then, it seems that Miozzo’s explanations of the anomic patients’ errors based on imageability are relevant, in one way or another, to our study.

If Andersen and Shirai’s *Aspect Hypothesis* and Collins’ (2007) findings are accurate, this has profound implications for our study. It means that our young learners should be more accurate with verbs belonging to the accomplishment and achievement categories. On the same account, our learners should perform more poorly with verbs belonging to the state and activity categories. All this, irrespective of experimental condition. In fact, in order to accurately assess the impact of the experimental conditions, a detailed analysis of the categories of verbs learners produce more errors in each condition is required. If no significant differences emerge, such errors should be attributed to the semantic categories, rather than to the experimental conditions.

In relation to our study, the most relatable of Ambridge’s (2010) findings and ensuing discussion is linked with the postulation that the English regular past tense forms might be memory-stored and, at the same time, generalizable analogically, based on their frequency (p. 1502). On the basis of such an account, the acquisition of English regular past tense forms can be explained by both analogies to stored regular forms and, at the same time, by rule-governance. Under the prism of our study, this might be interpreted as that in order to aid young EFL learners acquire the English
past tense, it is necessary that the input be modified to enable analogies between verb forms to be acquired and already existing verb forms. At the same time, the input needs to be modified in a way that aids our young learners form rules to aid them in the process of acquisition.

Zhuang’s (2010) time and space distinctions, cyclical interconnections between grammar elements, and the importance of vocabulary for implicit learning, are relevant for our purposes, as well. First of all, our learners should not only be taught how the English past tense localizes an event in the time frame, i.e. tense, but also how much space-time is encapsulated by an event in the past, i.e. aspect. In other words, the materials in our study should be carefully designed to capture both the time frame and the aspectual characteristics of the English past tense. Second, as to cyclical interconnections, this point reveals a fundamental drawback of our study, namely that it was not implemented repeatedly at regular intervals throughout the school year. In this way, the process of acquisition of the past tense by our young learners would have been examined in more precision. Finally, the cumulative store of lexical units of our learners was not taken into account and accurately assessed. Instead, their expected proficiency level was roughly estimated based on chronological age and prior educational experience. Notwithstanding, the results of our study should also be discussed taking into account Zuang’s argumentation.

Despite the fact that Mourssi’s (2012) classification does not directly correlate with the error patterns of the learners in our study, there are 2 types of errors that are relevant, Mourssi’s type 1 and type 5 errors / developmental stages. Taking into account that the learners in Mourssi’s study and the learners in our study had a different L1 background, it could be argued that, our young learners belonged to a developmentally earlier stage since the majority of errors they produced were non-
provision errors. As a matter of fact, according to our classification, Mourssi’s type 1 errors (including the provision of both base and present simple forms) were codified as 2 separate categories due to their large number. In other words, in our study base and present simple forms were classified as separate categories. In sum, if Mourssi’s and our own categorization of errors are accurate, then this could be regarded as constituting evidence that our young learners were true beginners. This means that our difficulties as to the initial classification of erroneous responses were, on the one hand, a problem, but, on the other hand, a helpful tool in assessing the real proficiency level of our young learners.

Lum and Kidd’s (2012) finding that their children performed better with regular verbs of high frequency compared to regular verbs of low frequency (p < .001), is of relevance to our purposes. All the more, since their tokens included a number of verbs also included in our study: (a) high frequency verbs such as needed, played, talked, watched, loved, jumped, and matched, and (b) the low frequency verb danced (p. 1004). Since Lum and Kidd’s subjects performed better on such high frequency verbs, the possibility that our subjects made many errors on such verbs might be a tentative sign of linguistic disabilities for our young learners.

Blom and Paradis’ (2012) findings are pertinent to our present purposes as well. The first is that there was variation between typically developing English L2 children and linguistically impaired English L2 children only as to regular past tense forms, but not as to irregular past tense forms (p. 289 – 290). The second was that the pronounced difference with past tense marking between these 2 groups of children involved verbs forming their past tense with the allomorph /d/. On the basis of such findings, Blom and Paradis concluded that linguistic impairment is linked with inefficient input processing skills (p. 292). However, the 2 authors had previously
commented that they could not verify that they could conclude that the past tense was a clinical marker for EFL learners (p. 290). In relation to our study, a detailed error analysis could be revealing as to whether the findings and ensuing discussion in Blom and Paradis (2013) are generalizable to our findings and general discussion.

7.7.3. The acquisition of the English past tense: summary

It is impossible to offer generalizations based on our study and the findings reported in the previous sections on a number of reasons. First, the research designs were markedly different. Second, these studies had different aims. Third, one of our basic aims was to compare implicit teaching methodologies whereas the basic aims of these studies included the acquisition of the English past tense regular and irregular forms. However, it became evident that the findings and ensuing discussion of such finding may offer tentative answers and conclusions to our questions and help address our aims as well. All in all, the inclusion of such findings and argumentation in our thesis renders our discussion more complete.

7.8. Discussion: Teaching Grammar in the L2

7.8.1. Discussion: Teaching grammar in the L2: introduction

Even though studies that include research designs that are similar to the present one as well as young children as participants are not common in the relevant SLA literature, in the present section, links with other similar research will be presented. The aim is (a) to cross-examine the findings of other similar research
designs, (b) to study how these findings were explained, and (c) in a way, see how
these explanations could also relate with the present study.

In other words, the present section includes a cross-examination of the
findings of the main study with relevant research findings. The discussion will be
limited to other studies utilizing in their research designs a combination of 2 out of the
3 following prerequisites: (a) typographical input enhancement, and/or input flooding,
and/or structured input, (b) young learners as participants, and (c) the English Past
Simple Tense as a target structure. Despite the fact that in chapter 3 the studies that
were reviewed did not include the aforementioned prerequisites, this decision is
motivated due to fact that cross-examining all those studies would be rather
complicated making it extremely hard to draw conclusions and reach valid
generalizations. The reason is that many of those studies utilized very different
research designs from the main study.

7.8.2. Discussion: Teaching grammar in the L2: the main study in relation to other
studies

7.8.2.1. Discussion: Teaching grammar in the L2: the main study in relation to other
studies: introduction

It needs to be emphasized that the vast majority of studies conducted within
the research paradigms of input enhancement and input flooding, and, especially
processing instruction and structured input, have utilized adult university level
students as participants. These studies are, in most of the cases, not cross-examined in
the present section. Throughout this section the age factor will be repeatedly referred
to since, as a variable, it presents, as already mentioned, a sharp contrast to other related studies.

Discussion in the following sub-sections is organized including, first, relevant studies within the textual input enhancement and input flooding frameworks, and, second, the processing instruction and structured input frameworks.

7.8.2.2. Discussion: Teaching grammar in the L2: the main study in relation to other studies: input enhancement & input flooding

To begin with, Trahey and White’s (1993) study involved 5th-grade French L1 children learning English in Canada. Despite reporting beneficial effects as to input flooding, the 2 authors also reported that their participants could not avoid the elimination of forms presenting similarities with their L1. This last issue was not examined in the main study. Trahey and White’s findings were only replicated in the main study as to the production of the target structure, not as to the interpretation of the target structure. The common feature between the main study and Trahey and White (1993) was input flooding. However, in the main study input flooding was conflated in the IF group with textual input enhancement in the form of bolding, consequently, any similarities between the 2 studies cannot be accurately substantiated.

White’s (1998) Group 2 did include textual enhancement and input flooding drawing on a participant pool including elementary school children. Much like the main study, White reported that input flooding resulted in variations in the way learners responded to the instructional materials. It seems then justified to assume that input flooding does procure beneficial learning outcomes but, at the same time, it is difficult to assess its exact impact on individual learners. Assuming that input
flooding is a highly implicit instructional treatment, this leads to the conclusion that learners respond to such implicit techniques in non-uniform ways. This means that this highly inductive approach may be suitable for some learners but not for all learners, some of whom might benefit from more explicit approaches. However, in the main study this was not a viable option for 8-year-old learners who lacked the skills to take part in and benefit from more explicit approaches. In all likelihood, it is ideal for young learners as a preparatory step before exposure to more explicit instructional techniques at older ages. Very similar to the main study, White alluded to the fact that input flooding present in the assessment materials may have been a causative variable for her results. This was a characteristic trait of our main study where the control group (only exposed to the assessment measures that did include input flooding) outperformed the 2 instructional treatment conditions.

Contrary to Lee (2007) who reported beneficial effects of textual enhancement for 11th-grade high-school children, the main study did not isolate textual enhancement in the form of bolding but conflated it with input flooding (IF group). The control group in the main study was not exposed to textual enhancement. Although it is easy to assume that textual enhancement may have left unaffected the IF group in the main study, nonetheless, definitive conclusions are unwarranted since textual enhancement was not applied in isolation. However, in the main study the control group that did not include textual enhancement outperformed the IF group. Assuming, like Lee (2007), who agrees with White (1998), that textual enhancement is more explicit than input flooding, and taking into account the performance of the control group, may lead to an extreme, but viable, conclusion, i.e. that textual enhancement may have indeed had a distracting role as to the interpretation of the target structure.
This, however, contrasts with Simard’s (2009) proposal that textual enhancement coupled with input processing may be beneficial for some target forms. Simard’s pupils in secondary one grade were exposed to different textual enhancement techniques. In other words, Simard isolated different textual enhancement techniques as variables. However, she did not find any beneficial effects for bolding, the textual enhancement technique utilized in the main study. Simard made explicit reference to the fact that capitalization and the 3-cues combination (3 textual enhancement types combined) procured the best learning outcomes compared to the rest of the textual enhancement techniques used. This might be interpreted in the framework of the main study as follows: either bolding is not as beneficial as capitalization (one of the most beneficial techniques reported by Simard), or that individual learners or learner groups profit more from different textual enhancement techniques.

Sarboland (2012), on the other hand, reported beneficial effects as to the acquisition of the English Past Simple Tense by the textual enhancement techniques of underlining and bolding. The participants in Sarboland’s study included both adolescents and adults. Taking into account the main study with young 8-year-old pupils as participants and Simard’s (2009) study where the participants were secondary school pupils, it might be tentatively proposed that learners of different ages profit the most from different textual enhancement techniques, i.e. for adolescents and adults bolding may be beneficial, whereas for younger learners it might operate as a distracting factor. Another variable may be the learners’ L1, but neither the present study, nor Simard’s (2009) or Sarboland’s (2012) studies can substantiate such a claim. Neither Simard (2009), nor Sarboland (2012) reported input flooding as a variable in their research designs.
Textual enhancement, as to the acquisition of the English Past Simple Tense, was also investigated by Nahavandi and Mukundan (2013). The 2 authors combined textual input enhancement with rule presentation in 1 of their 3 groups and reported statistically significant beneficial effects in comparison to a control group. From the perspective of the main study 2 issues merit attention. The 1st issue concerns the age factor. Nahavandi and Mukundan included adults as participants. This means that probably these adults, even though at the elementary level as to EFL learning, were accustomed to rule presentations, or were at least capable of profiting from such explanations. The young learners in the main study would not, in all likelihood, profit the same because they were probably unaccustomed to such explicit explanations and because at such a young age explicit rule presentations can be hardly understood. The 2nd issue is related to the fact that in the main study textual enhancement for the IF experimental groups was conflated with another variable, i.e. input flooding. This research design seemed more appropriate for the 8-year-old children in the main study. To the contrary, Nahavandi and Mukundan isolated and worked on textual enhancement exclusively. In fact, this final issue, i.e. the unique conflation of variables in the main study, renders comparisons between the main study and other relevant studies almost impossible since most often the respective research designs differ markedly.

This final issue also pertains to the study of Rikhtegar and Gholami (2015) in relation to the main study. These 2 authors investigated the English Past Simple Tense in the framework of input flooding and rule explanation. The argumentation is that similar to the Nahavandi and Mukundan (2013) study; that in the main study input flooding was not used in isolation. However, it was operant in the assessment measures and the control group seems to have profited from it despite the fact that it
did not receive any kind of rule explanation. Rightegar and Gholami’s (2015) participants were 11 to 14 years old. Once again, they could profit, and indeed did profit, from rule explanations, an option not so applicable to the participants in the main study due to their age.

LaBrozzi (2014) investigated the relative effects of various textual enhancement types in a research design using adult learners at a university in the U.S.A. The results revealed a hierarchy of effectiveness of textual enhancement types in order from more beneficial to less beneficial: (a) changes in font size, (b) capitalization, (c) bolding and italics, and (d) changes in font and underlining. It seems that much like the young learners in Simard’s (2009) study, the adult learners did not respond so well to the textual input enhancement (TIE) technique of bolding.

La Brozzi’s (2014) and Simard’s (2009) findings may then explain why the TIE technique utilized in the main study, i.e. bolding, failed to procure the desired results in comparison to the control group who received an input flood through the tests, without any enhancement such as bolding whatsoever.

7.8.2.3. Teaching grammar in the L2: the main study in relation to other studies: structured input & processing instruction

The research design of the main study can be originally traced to VanPatten and Oikennon’s (1996) study, and then to Farley (2004b). Both these studies examined structured input activities as the causative variable of processing instruction. VanPatten and Oikennon (1996) reported that structured input activities, and not the explanation phase of PI, were the causative variable for the enhanced learning outcomes. Farley (2004b) reported findings in favour of processing instruction as compared to structured input only. In the main study, the explanation
phase of processing instruction was omitted and only structured input was utilized. As already mentioned, young learners in the main study were at an age when rule explanations are not so suitable. Moreover, Farley’s SI group failed to procure enhanced learning outcomes compared with the PI group. This resembles the findings in our main study where the SI group procured the least beneficial effects. Therefore, it can be assumed that rule explanation might be a necessary component of PI since SI activities on their own are incapable of exhibiting enhanced learning outcomes, resembling Farley’s (2004b) findings. In fact, it could be argued that, on the whole, PI and SI are more suitable for older learners, especially taking into account the fact that PI and SI involve the altering of erroneous processing strategies. In all likelihood, younger learners profit more from less explicit instructional treatments like input flooding.

Erlam (2003) also included an SI group in the research design. However, her findings are contrary to the findings of the main study. More specifically, Erlam’s SI group made greater gains than the C group, whereas in the main study the C group made greater gains than the SI group. Moreover, Erlam’s output-based group outperformed the SI group. Consequently, in Erlam’s study the SI group lagged in respect to another experimental instructional treatment. Erlam’s participants were not adults (like Farley, 2004b), but secondary school students.

As a general remark, then, it seems that SI, on its own, has generated mixed results in the relevant literature and there are a number of studies that have reported relevant non-beneficial effects for different sample populations: Farley (2004b) included university students, Erlam (2003) included secondary school students, and the main study included 3rd-grade Primary School pupils.
Benati (2005) reported a study that, similar to the main study, investigated the acquisition of the English Past Simple Tense. The participant pool included Greek L1 EFL learners aged 12-13 years old. There are 3 major differences between the main study and Benati’s study. The 1st difference concerns the participants. The participants in the main study were 8 years old, i.e. 4 to 5 years younger than the participants in Benati’s study. The 2nd difference is that Benati utilized a PI group whereas in the main study an SI group was incorporated. The 3rd difference is related with the fact that the other 2 experimental conditions in Benati’s study were traditional instruction (TI) and meaning-based output instruction (MOI) and no control group was reported, while the main study included an input flood (IF) group, as well as a control group. In other words, despite investigating the same target form, valid cross-examinations between Benati (2005) and the main study are not substantiated. Finally, even though this author reported that SI is the reason behind improved PI outcomes, the main study failed to substantiate this claim. Not only that, but it was the control non-treatment condition that disproved this claim. In other words, despite superficial similarities, the main study and the Benati (2005) study differ in fundamental ways.

On the other hand, despite the fact that Marsden (2006) and the main study do not share many common elements, there is a notable similarity. More specifically, the author reported that her “test-only class” (p. 541) (something like a control group although not overtly stated as control by Marsden) exhibited enhanced learning outcomes in all measures. This mirrors the findings of the main study, even though Marsden utilized a PI group and the present study utilized an SI group. Marsden attributed the enhanced learning outcomes of the test-only class to the fact that these learners were accustomed to grammar-oriented tasks and testing materials. The
present study attributes similar control group findings to a host of other factors put forth in section 7.6.

Lee and Benati (2007), in their book, reported on a number of studies that included learner groups exposed to both SI and IE, which they labelled SIAE. Lee and Benati concluded in 1 of their studies that

“(u)nenhanced structured input was equally as effective as enhanced structured input practice in helping learners to interpret and produce accurately” (Lee and Benati, 2007, p. 109, parenthesis added)

In other words, Lee and Benati considered SI as the causative variable of the enhanced learning outcomes and not input enhancement (IE). This can be considered as evidence that IE does not affect learning, or to be more accurate, does not in combination enhance the gains promoted by SI, i.e. it does not make any measurable contribution to activities which have been structured according to IP and PI guidelines. It should be underscored that Lee and Benati’s studies that included SIAE involved university level students. There are 2 fundamental differences between the studies in Lee and Benati (2007) and the present study. The 1st difference is that the inclusion of a control group was not usually reported by the 2 authors and the 2nd difference is that input flooding was not included as a variable. For these reasons, valid generalizations are not substantiated.

Qin (2008) used a research design that included both PI and dictogloss (DG). Thirteen- to 15-year-old secondary school children participated. Qin reported enhanced learning outcomes for both groups. This study is 1 of the few cases where young learners were exposed to PI. What is more, it seems that they also profited significantly from such exposure, contrary to the main study where the SI group not
only failed to outperform the other 2 groups post-experimentally, but exhibited some kind of detrimental effect as to the interpretation of the target structure.

Another relevant study including 6th grade primary school pupils and investigating the acquisition of the regular English Past Simple Tense –ed was Marsden and Chen (2011). It should be noted that in this study the authors investigated the effects of the 2 components of SI, i.e. referential and affective activities. From the perspective of the main study, their finding that affective activities did not procure beneficial learning outcomes is of importance. The reason is that in the main study the participants were exposed to 2 referential and 2 affective activities. If Marsden and Chen’s postulation is accurate, then this could imply that the participants in the main study did not receive adequate or appropriate exposure to the target form in the form of referential activities. In other words, to acquire the regular English Past Simple Tense –ed young learners may need more abundant exposure to this form through a large number and variety of referential activities. Maybe, then, insufficient exposure to the target form may be the causative variable behind the performance of the SI group in the main study.

Nonetheless, Marsden and Chen (2011) seem to have downplayed the fact that affective activities are the element of PI which is learner-friendly, even though they do mention this particular point (p. 1062). In other words, they seem to imply that learner friendliness in the materials does not enhance learning outcomes. This interpretation of their findings underestimates the fact that affective factors incorporated in learning agendas are a prerequisite for learning, even though they are difficult to accurately report and cannot easily procure statistically significant results as a separate variable.
7.8.2.4. Discussion: Teaching grammar in the L2: the main study in relation to other studies: summary

All in all, after the cross-examination of the aforementioned studies with the present main study, the factor that mostly seems to differentiate the main study from the other studies concerns the age of the participants. Also, it seems likely that the differences as to the results may be attributed to other factors such as different research designs. Nonetheless, the fact that the research design in the main study was necessitated mainly by the young age of the participants renders the age factor as the variable that differentiates fundamentally the main study from other relevant ones. In other words, it is hard to use the explanations other researchers have used to explain the present findings.

7.8.3. Teaching grammar in the L2: summary

Lee and Huang (2008) underscored the issue of proficiency levels of the subjects in the studies reviewed in their meta-analytic review. More specifically, they noted the overall tendency of research within the visual input enhancement framework to include intermediate level learners. Lee and Huang attributed this tendency to the assumption that beginners might not possess the necessary reading skills to engage in visual input enhancement research, whereas advanced learners were probably already very competent as to the target structures investigated. This is assumedly a gap in the literature, since prior knowledge – irrespective of the means of assessment – is a confounding variable that might distort the results. In this respect, Lee and Huang noted that studies including learners with high levels of prior knowledge need to be interpreted with caution.
Dekeyser and Botana (2015), after conducting a review of the theoretical and research literature within the framework of PI acknowledged the fact that sample populations are most of the times young adults, participants in learning programmes in tertiary-level institutions in the U.S.A. (p. 291 – 292). Shintani (2015) also conducted a meta-analysis including PI studies and production-based instruction (PB) studies. In fact, Shintani’s meta-analysis included only studies with participants who were either adults or adolescents. Generally speaking, in these studies adults performed better than adolescents. As Dekeyser and Botana (2015) pointed out [and this also relates to Shintani (2015)], “this clearly limits the generalizability of the findings to younger learners” (p. 300).

The cross-examination of relevant studies within the frameworks of textual input enhancement and input flooding, and processing instruction and structured input reveals that the age factor limits direct comparisons since the majority of studies have not used young Primary School learners as participants. This is especially pronounced within the frameworks of processing instruction and structured input framework where, with notable exceptions, the studies have utilized young university-level adults as participants. A relevant difference between the main study and studies within the textual input enhancement and input flooding frameworks concerns the fact that these teaching methodologies have been investigated mainly with intermediate level learners in sharp contrast with the main study where the learners were late beginners.

7.9. Discussion: Teaching the regular English Past Simple Tense -ed to Young Greek L1 EFL learners

7.9.1. Discussion: Teaching the regular English Past Simple Tense -ed to Young Greek L1 EFL learners: introduction
The regular English Past Simple Tense -ed is explicitly taught for the first time as a grammatical phenomenon in the 5th grade of Primary School in Greece. The State-appointed English coursebook (Αγγλικά Ε' Δημοτικού, English 5th Grade, Kolovou & Kraniotou, 2010) is obligatorily taught in all Greek Primary Schools. The regular Past Simple Tense –ed is taught in Unit 7, Lesson 1 (affirmative) and Lesson 2 (negatives and questions) (p. 85-96).

7.9.2. Discussion: Teaching the regular English Past Simple Tense -ed to Young Greek L1 EFL learners: curricular issues

In the relevant section of the English 5th Grade Teacher’s Book (Kolovou & Kraniotou, 2009) containing methodological tips regarding the teaching of grammar, a number of guidelines are presented. According to the Teacher’s Book, grammatical forms are deconstructed (form – meaning – use) after exposure to the target form to enhance awareness and noticing. The grammar is either (a) situationally presented, or (b) structured to be acquired inductively. Practice takes the form of (a) drilling to produce controlled output and (b) more learner-friendly grammar activities (quizzes, memory tests, picture dictation) aiming at free production (p. 12 - 13).

In the relevant ‘Grammar Focus’ section(s) of the Pupil’s Book, the grammar structures are input flooded and visually enhanced (in bold form). The regular Past Simple Tense is presented in the Pupil’s Book Grammar Focus sections (Kolovou & Kraniotou, 2010, p. 88 and 92). In the Discover Grammar Appendix at the end of the Pupil’s Book, form, meaning, and use are explicitly presented and rule-stated, and declension tables are included (p. 158 - 159).

Throughout Unit 7 (like in all other units), grammar is highly contextualized, all four skills are practiced extensively, and inductive learning is mainly promoted
through interesting and appropriate thematic settings. A variety of organisations of study are promoted: individual, pair-work, group-work, whole-class participation. Portfolio sections, cross-curricular projects, and a self-assessment test at the end of each unit are included, as well.

In the present study, the possibility of enhancing learning outcomes concerning the English regular Past Simple Tense -ed within the Greek State Primary School system was examined. There were two possible options. The first option was to enhance exposure for the 10-year-old learners of the 5th grade. The second was to examine the option of exposure - teaching to even younger pupils.

In the former case, the overall Greek State Primary School curriculum would probably end up assigning more classroom time to the teaching of the English language as a school subject than at present. In the Greek Primary School curriculum English as a school subject is currently being granted more classroom time than all other subjects (Physical Education being an exception), coming second only to the Greek language as a subject.

The present study examined the second option, i.e. that of exposing younger learners to a grammatical structure like the regular Past Simple Tense -ed. In the pilot study, 3rd grade pupils were selected. On the basis of the findings of the pilot study, the decision was taken to try and adapt more age-appropriately the materials to such learners. Consequently, in the present study, 3rd grade pupils were selected, as well. The option of selecting 1st or 2nd grade pupils was rejected, because these learners lack the skills required to participate in the structured input (SI) and input flood (IF) instructional treatments and the interpretation and production assessment measures as operationalized in our main study.
A similar consideration was reported by Han and Liu (2013) who stressed the need of using as participants and assessing what they termed *ab initio* learners, i.e. genuine beginners … who, therefore have had zero to little experience of the target language” (p. 145). The two authors attributed the relatively high levels of comprehension by both groups of learners in their study on prior experience and knowledge of the world (p. 160). The same issue was raised by Godfroid (2015) who resorted to utilizing and experimenting on semi-artificial L2 German in an attempt to control learners’ prior knowledge of the targeted forms as well as their overall exposure and input of the target language (p. 6). This is a very crucial issue since prior knowledge of targeted forms may contaminate the results borne out by measures of learning outcomes consisting impossible the task of whether reported learning outcomes are to be attributed to (a) instructional treatment conditions, (b) previous knowledge of target forms, or (c) both. Coupled with the fact that previous exposure to a target language through formal instruction may unconsciously raise learners’ awareness of language properties including target forms absent in the input, raises serious doubts as to whether the reported beneficial effects of instruction can be attributed to instructional treatment conditions per se. The matter is further complicated taking into account that adult L2 learners are equipped with real world knowledge and cognitive schemata that implicitly, but substantially, affect their performance as (L2) learners.

To the researcher aware of these difficulties, three options seem profitable: (a) to utilize artificial or semi-artificial target languages, (b) to include young beginner-level natural L2 learners as sample populations, or (c) both. Option c includes the risk of learners not being able to participate due to inefficient learning skills. Option a needs to cautiously handle the issue of what constitutes an artificial or semi-artificial
language, i.e. whether it conforms to Universal Grammar (UG) principles and parameters. This makes option b as the most viable option that compromises both options a and c while, at the same time, maintaining theoretical, practical, and ecological validity to any research design.

7.9.3. Discussion: Teaching the regular English Past Simple Tense –ed to Young Greek L1 EFL learners: summary

Despite the fact that the regular English Past Simple Tense –ed target structure is explicitly taught at the 3rd trimester of the 5th grade, the results of the present study showed that there are ways of preparing this explicit teaching, even at the end of the 3rd grade. It seems that even though these learners were exposed to implicit instructional settings, nonetheless, they managed to exhibit signs of acquisition. Moreover, much like regular classroom settings where learners of different proficiency levels exist, all the participants were included in the statistical analyses and no cut-off score was used. This closely resembles real classrooms.

Even though the results of the present study seem contrary to curricular decisions to explicitly expose Greek Primary School learners to the target structure in the 5th grade, in fact, they are not. Instead, the findings could be informative as to the exposure to the target structure at the 4th or 3rd grade using implicit or implicit-like instructional conditions. In other words, implicit exposure to the English regular Past Simple Tense –ed could act as a precursor, i.e. an aid, to explicit exposure at the 5th grade. More specifically, in the 3rd grade, pupils could be exposed to the target structure using the novel implicit techniques of input enhancement and input flooding, even structured input, and by the 5th grade they would have the adequate implicit cognitive schemata to tackle the explicit load of the target structure. Nonetheless, it needs to be emphasized that age-appropriate materials need to be designed and that
the treatments need to be as unobtrusive as possible. Ideally, the young learners should be unaware that they are being taught. Instead, they should be called to take part in novel game-like activities.

7.10. Discussion: Implications for Language Teaching

There are 3 major implications for language teaching that can be drawn on the basis of the findings of the main study in the present thesis. The first is that teaching can tamper with developmental sequences. In our case 3rd-grade 8-year-old EFL learners exhibited partial mastery of a grammatical phenomenon they were supposed to be exposed at the age of 10 in the 5th grade. The second implication is that there is broad scope for attempting to teach our young EFL learners structures that may be far beyond their current proficiency level. In fact, aspects of language can be taught to all / any learner(s) irrespective of proficiency level if organized, presented and taught according to mental and cognitive capacities of learners (Bruner, 1966). The third implication is related with discovery learning, i.e. there is no need for explicit explanation; the thing that counts most is that materials need to be organized to aid learners process target forms on their own (Bruner, 1966).

7.11. Discussion: Implications for Teaching Young Learners

In the same vein, based on our study, 3 implications concerning the teaching of young EFL learners can be drawn. The first implication concerns the need to adapt the materials to the learners and not vice-versa, i.e. the learners to the materials. The main study has shown that it is possible to teach young learners target forms beyond
their current proficiency learners by designing age-appropriate materials. The second
implication relates with the fact that for young learners implicit methodologies and
teaching methodologies can be a beneficial, if not optimal, way of teaching. In this
way, such learners are not exposed to (meta)linguistic explanations that they may not
be capable of assimilating. Instead, they are free to process the material on their own
and at their own speed. The third implication is that, for this age group, simple
abundance of the tokens in the input suffices for acquisition following the tenets of
naturalistic L1 learning.

7.12. Discussion: Implications for Implicit Learning

Based on our present findings a number of issues are raised concerning
implicit learning methods. The first issue is related with the fact implicit learning is
susceptible to individual variation. This is an expected outcome due to the element of
freedom incorporated in such methodologies. The standard deviation values reported
are valid proof for this state of affairs. The second issue will be repeated as it has been
raised in other parts as well. It concerns the fact that implicit methodologies are
appropriate for teaching young learners. Although explicit instruction was not
included as an instructional condition in the main study, still, the results do point to
the direction that our young EFL learners can and do learn implicitly. The third issue
is related with the tentative assumption that implicit methodologies procure better
results as to production than as to interpretation. This might have also been due to
inefficiencies of our research design, but it is, nonetheless, substantiated by our
results. The final issue is related with the crucial decision of the amount and type of
exposure necessary to procure the desired learning outcomes by our implicit
methodologies. Despite the fact that we were obliged to accurately balance the amount of input provided to each experimental conditions, it remains a fact that this might have disregarded the fact that, for example, learners in the IF group might have required more tokens in the input to procure the desired results. In other words, implicit learning methodologies may require not only a qualitative, but also a quantitative differentiation.

7.13. Limitations

This section is a brief exposition of the limitations that, undoubtedly, exist in the main study. The reason for presenting these is twofold: (a) to facilitate the assessment of the outcomes and (b) to help others interested in carrying out similar research. Therefore, it can be viewed as some indication of what other researchers should attend to, based on what I should have paid more attention to.

To begin with, a crucial element in any relevant research design concerns the amount of input the learners are to be exposed to. Despite the fact that in the main study the learners who received the less amount of input, i.e. the control group, outperformed the other 2 groups, i.e. SI and IF, the results might have been different if the amount of exposure to the target structure offered to the other 2 groups had been more abundant. After cross-examining 2 SI groups of learners, one receiving 16 SI activities, and another one receiving 8 SI activities, Li reported that the former group outperformed the latter one on a number of measures. Li attributed this enhanced performance to the fact that the learners in the former group had received a larger number of activities (Li, 2012, p. 425 – 427). Maybe for the young learners who participated in the main study more structured input and input flood activities would
have procured more beneficial results. Despite the fact that the amount of input offered was balanced due to the very small number of available tokens, maybe the learners should have been taught more instances of regular English Past Simple –ed verbs, through e.g. the teaching of a small list of verbs pre-experimentally. Of course, there are limits to the amount of vocabulary such young learners can acquire, but the limitation remains.

The 2nd limitation is related with the assessment measures employed. More specifically, spontaneous oral speech production measures could have been incorporated. Such tests were not administered due to 2 reasons: (a) the young age of the learners and (b) time constraints. The young age of the learners meant that these learners lacked the appropriate speaking skills to participate in such tasks. On the other hand, one-to-one oral interviews would need ample time not available due to the fact that both the pilot study and the present study were administered during the final days of the school year.

A 3rd limitation (related to the 2nd one) was that long-term instructional effects were not assessed in the absence of delayed posttests. This was, once again, due to the fact that the main study was administered at the end of the school year. However, this remains a serious limitation since there is a likelihood that the learners in the SI and IF instructional treatments might have been able to restructure their interlanguages (ILs) in the meantime, thusly allowing their input to become intake. Maybe the crucial element was that SI and IF should have been granted a lengthier time span to exhibit enhanced learning outcomes.

A 4th limitation was that more information about the sample population was not requested. In other words, I could have asked for more information about the learners as to their learning methods, performance in tests, previous exposure to other
(similar) target forms, exposure to English, to name but a few possible avenues. This would have enabled a more accurate assessment of the learning outcomes as well as cross-examinations with other similar sample populations. In this respect, it is difficult to generalize the findings of the main study to other sample populations, despite the fact that the statistical analyses reporting statistical significant results are supposed to tackle such a problem.

Finally, the sample size was quite small and the number of participants in each group unequal. Li also reported small sample size as a limitation (p. 429 -430). This means that however accurate as to generalizations the statistical analyses may have been, it is a fact that, in order to offer accurate results and research findings, the study needs to be replicated including a much larger number of participants more evenly distributed among groups. These are crucial issues. They could have been overcome if the study had been carried out in more schools or including other target groups.


Turning now to areas of further investigation, a number of suggestions to those interested can be offered. First, a most challenging task would be to attempt and conduct similar studies with even younger learners. Despite the fact that many of the limitations of the main study are linked, in one way or another, to the age factor, the challenge remains, i.e. to expose even younger learners to similar conditions in similar research designs. In all cases, the main study is valid proof that instructional treatments thus far mainly involving adult sample populations, can be, and should be, addressed to younger learners. This is a very lucrative prospect in EFL, especially since the teaching of English is of primary importance in many countries.
Moreover, there is an ever-growing need to provide longitudinal data as to the efficacy of the instructional treatments employed in the main study, especially from an educational point of view. Maybe input enhancement, input flooding, and structured input can be scheduled as parts of the regular ELT curricula and cease being the objects of isolated research agendas not directly linked with actual teaching practice. It is high time that the present and similar research finding be adapted to mainstream teaching practice, after being studied in longitudinal research designs.

Finally, similar age sample populations from other L1 backgrounds could be investigated in order to enable more accurate cross-examinations with other L1s. This would offer valuable insights and at the same time permit to incorporate findings from target populations from other countries, cultures, and linguistic backgrounds. In this respect, language might become a bridge with other civilizations, instead of a barrier. This applies especially for English which has acquired the status of a lingua franca, in other words a primary vehicle of communication around the world.
GENERAL CONCLUSION

There are but a few studies in the relevant literature combining (a) the implicit methodologies of SI and IF & TIE, (b) the English Past Simple Tense –ed as a target structure, and (c) young 8-year-old learners as participants. Our main contribution lies on the fact that it was proven that target forms beyond young learners’ proficiency level are amenable to instruction if the materials are appropriately designed and implemented to match the age and the cognitive level of such learners. Our study paves the way for conducting similar research with young learners as participants.
GREEK


Καψάλης, Γ., Α. (2003). Παιδαγωγική Ψυχολογία. Θεσσαλονίκη: Εκδοτικός Οίκος Αδελφών Κυριακίδη Α. Ε.

ENGLISH


Multilingual Matters.


