EXPERIENCES OF REMEMBERING AND KNOWING IN SLA, L2 DEVELOPMENT, AND TEXT COMPREHENSION: A STUDY OF LEVELS OF AWARENESS, TYPE OF GLOSSING, AND TYPE OF LINGUISTIC ITEM

A Dissertation submitted to the Faculty of the Graduate School of Arts and Sciences of Georgetown University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Spanish and Portuguese

By

Ana María Martínez Fernández, M.S.

Washington, DC August 9, 2010
EXPERIENCES OF REMEMBERING AND KNOWING IN SLA, L2 DEVELOPMENT, AND TEXT COMPREHENSION: A STUDY OF LEVELS OF AWARENESS, TYPE OF GLOSSING, AND TYPE OF LINGUISTIC ITEM

Ana María Martínez Fernández, M.S.
Thesis Advisor: Ronald P. Leow, Ph.D.

ABSTRACT

Attentional models in SLA (e.g., Schmidt, 1990; Robinson, 1995b) propose a crucial role for awareness in L2 learning, and suggest that awareness allows for encoding in episodic memory. In cognitive psychology, Tulving (1983) introduced the notions of episodic memory, which involves recollection of contextual details (i.e., remembering), and semantic memory, characterized by a sense of familiarity (i.e., knowing). Studies indicate a relationship between awareness and remembering (cf. Gardiner, 2008). To date, however, memory experiences have not been empirically investigated in SLA.

SLA researchers have been more concerned with the effectiveness of pedagogical techniques to promote learners' awareness. The present study investigates memory experiences and awareness in different glossing conditions. Studies addressing effects of type of glossing on learning and text comprehension have yielded mixed results, and overall, they do not account for a type of item effect. Moreover, although glossing is premised on learners’ attention to the glosses, few studies have measured attention and awareness (e.g., by employing concurrent verbal reports).

Within an attentional framework, this study investigates remembering and knowing in SLA in relation to L2 development, levels of awareness, type of glossing in a
reading comprehension task, and type of linguistic item. Intermediate learners of Spanish read two texts containing lexical and grammatical items under one of five conditions that differed in (a) type of exposure (gloss, gloss embedded in a fill-in task, no gloss), and (b) whether or not they were asked to think aloud while reading. Participants completed immediate and one-week delayed posttests, including text comprehension questionnaires, the remember-know task, and word meaning and grammar production and recognition tests.

The results of the study provide evidence for the relationship between remembering and L2 development. Furthermore, the study supports (a) a correlation of awareness with remembering of grammatical items and L2 development overall, and (b) a beneficial effect of glosses on vocabulary noticing and learning, and text comprehension. However, no evidence for the effect of glossing on memory experiences, awareness or learning of grammatical items was found. Finally, findings suggest that vocabulary is further processed into episodic and semantic memory to a greater extent than grammar.
ACKNOWLEDGEMENTS

There are many people whom I would like to thank for their contribution to this dissertation and their support since I started the Ph.D. program. The idea for this dissertation was developed during a seminar on awareness and L2 learning that Dr. Ron Leow offered four years ago. I thank Ron for all that I have learned from him since then, both as a student and as a person, for his unconditional support when I decided to move back to Spain two years ago, and for making it so easy to work in the long distance. I also thank him for his tremendous effort to be fair with all students, for his sense of humor, for e-mailing me to support Spain a few minutes before winning the World Cup final, for the bottle of champagne he brought to my defense, for being critical with the work and warm with the people, and for inspiring me to do always my best. Thank you, Ron.

I would also like to thank the members of my committee: Dr. Andrea Révész and Dr. Rusan Chen. Thanks to Dr. Révész particularly for her insightful comments and advice on the research design, which was greatly enhanced by her guidance. Thanks to Dr. Chen for his help with the statistical aspects of this dissertation. Special thanks to my advisor and the members of my committee for being available during their holidays and allowing me to defend my proposal and dissertation when it worked best for me.

Very special thanks to my friend and colleague Miguel Ángel Novella for his priceless help with data collection while I was in Barcelona and when I came to DC. I also thank Miguel for his enormous generosity, for hosting me and picking me up at the airport several times, for forcing me to buy a pair of snow boots when I was trying to collect data in the middle of a historic snowstorm, and for his constant friendship.
Thanks to Peter Janssens and Susan Pennestri for their technical support and for allowing me to use the language laboratory for data collection. Thanks to the program directors Ron Leow and Cristina Sanz, the coordinators Julio Torres and Missy Baralt, and all the instructors who allowed me to collect data from their students. Thanks to many colleagues who have helped me along the way: Mika Hama, Missy Baralt, Laura Weiss, Claudia Guidi, Paco Fernández-Rubiera, Carolina Rodríguez.

Many thanks to Grant Armstrong and Ana Armstrong for hosting me several times. I especially thank Grant for being a friend I could always trust, and for all the good moments we have shared in and outside the Big Hunt during our graduate studies. I also thank my very first friends in the US, Alex, David, Gon, and Javi, who made my arrival to this country unforgettable.

There is also a long list of people in Spain to whom I owe my gratitude. I thank Dr. Alfonso Pitarque for kindly answering my questions on methodology and offering his help. I thank my friends and colleagues in Barcelona, Anna, Darcy, Elena, Helena, Laura, Lizzie, Lucía, Rebeca, and Rosalía (in strict alphabetical order), who have seen me, day by day, going through the final stages of my dissertation, and who have supported me when I needed it the most. I also thank my dear friends Carlota, Caro, Marisa, Rossa, Coni, Laura, Diego, Gabi, and Sonsoles (in strict randomized order), who have always been near despite the long distance. I thank Detlef Hoffman and Cristina Nadal, for their crucial support in my own experiences of memory during the last two years. I especially thank my mother and my brother Fermín for traveling all the way to DC to come to my defense and share that important moment with me. Thank you to all of you.
DEDICATION

To my grandmother, Ángela Fernández López (1917-1991), and to my mother, Ana María Fernández Fernández.
### TABLE OF CONTENTS

**INTRODUCTION** ....................................................................................................................... 1

**STATEMENT OF THE PROBLEM** .................................................................................................... 1

**DEFINITION OF TERMS** ............................................................................................................... 4

**CHAPTER 1: REVIEW OF RELATED LITERATURE** ................................................................. 10

**THEORETICAL FOUNDATION** ...................................................................................................... 10

Models of Recognition Memory Experiences in Cognitive Psychology: Remembering and Knowing ........................................................................................................................................... 11

- Tulving’s model of episodic and semantic memory systems .............................................. 11
- Dual-process models .............................................................................................................. 17
- Single-process models .......................................................................................................... 20
- Recent developments ............................................................................................................. 22
- Conclusion .............................................................................................................................. 24

Models of Memory and/or Awareness in SLA ........................................................................... 25

- Schmidt’s Noticing Hypothesis .......................................................................................... 25
- Robinson’s Model of the Relationship Between Attention and Memory ...................... 27
- N. Ellis’s Model of Memory for Language ......................................................................... 28
- Ullman’s Model of Declarative/Procedural Memory .......................................................... 31
- Conclusion .............................................................................................................................. 32

**EMPIRICAL STUDIES ON THE RELATIONSHIP BETWEEN MEMORY EXPERIENCES AND AWARENESS** ........................................................................................................... 34

Empirical Studies in Cognitive Psychology ............................................................................... 34

Empirical Studies in SLA .......................................................................................................... 44
LIST OF FIGURES

Figure 1. Procedure ......................................................................................................... 159

Figure 2. Meaning Recognition of Target Lexical Items .................................................. 195

Figure 3. Meaning Recognition of Lexical Items Overall .............................................. 196

Figure 4. Grammar Recognition: Old Items ................................................................. 201

Figure 5. Grammar Recognition: New Items ................................................................. 201

Figure 6. Meaning Production of Target Lexical Items .................................................. 206

Figure 7. Grammar Production: Old Items ..................................................................... 212

Figure 8. Grammar Production: New Items .................................................................. 213

Figure 9. Text Comprehension of Global and Local Ideas on the Lexical Item Text .... 215

Figure 10. Reported Memory Experiences by Type of Linguistic Item ......................... 217

Figure 11. Reported Awareness of Vocabulary by Reported Level of Awareness of
Grammar ....................................................................................................................... 220

Figure 12. Recognition by Type of Linguistic Item (Percentage) .................................. 224

Figure 13. Production by Type of Linguistic Item (Percentage) ..................................... 227
LIST OF TABLES

Table 1. Remember-know studies addressing effects of LOP on recognition memory of L1 words ................................................................. 37
Table 2. Studies addressing the generation effect on memory experiences .................. 38
Table 3. Studies addressing the effect of undivided vs. divided attention during encoding on memory experiences ........................................................................................................ 39
Table 4. Studies addressing the effect of encoding words vs. non-words on memory experiences ............................................................................................................................................. 40
Table 5. Variables by Research Question .................................................................. 125
Table 6. Means, SD, and SE for Text Comprehension Pretests by Group .................. 161
Table 7. Means, SD, and SE for Grammar Recognition Pretests for Two Types of Item by Group ............................................................................................................................ 162
Table 8. Means, SD, and SE for Grammar Production Pretests for Two Types of Item by Group ............................................................................................................................ 163
Table 9. Means, SD, and SE for Fill-in Task Performance on Lexical and Grammatical Items by Think-aloud Group ........................................................................................................ 164
Table 10. Means, SD, and SE for Text Comprehension of Two Texts by Think-aloud Group ........................................................................................................................................... 165
Table 11. Means, SD, and SE for Reported Memory Experiences of Lexical and Grammatical Items by Think-aloud Group ............................................................... 166
Table 12. Means, SD, an SE for Word Meaning and Grammar Recognition by Time and Think-aloud Group ................................................................................................................ 168
Table 13. Means, SD, an SE for Word Meaning and Grammar Production Tests by Time and Think-aloud Group

Table 14. Means, SD, and SE for Time on Task (minutes) by Group and Text

Table 15. Means, SD, and SE for Time on Task (minutes) by Type of Glossing and Text

Table 16. Means, SD, and SE for Global Comprehension by Text

Table 17. Tests Reliability

Table 18. Correlation Analyses of Reported Memory Experiences and Recognition

Table 19. Correlation Analyses of Reported Memory Experiences and Production

Table 20. Correlation Analysis of Reported Levels of Awareness and Reported Memory Experiences

Table 21. Correlation Analysis of Reported Levels of Awareness and Recognition

Table 22. Correlation Analysis of Reported Levels of Awareness and Production

Table 23. Means, SD, and SE for Reported Memory Experiences of Lexical Items by Type of Glossing

Table 24. Mean, SD, and SE for Reported Memory Experiences of Grammatical Items by Type of Glossing

Table 25. Means, SD, and SE for Reported Awareness of Lexical Items

Table 26. Means, SD, and SE for Reported Awareness of Grammatical Items as Measured by Online Procedures

Table 27. Means, SD, and SE for Reported Awareness of Grammatical Items as Measured by Online and/or Offline Procedures
Table 28. ANOVA on Meaning Recognition of Target Lexical Items by Time and Group
.................................................................................................................................................. 190
Table 29. Means, SD, and SE for Meaning Recognition of Target Lexical Items by Type of Glossing and Time................................................................................................................. 191
Table 30. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [-I] Group by Time: Pretest versus Immediate Posttest...................................................... 192
Table 31. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [-I] Group by Time: Pretest versus Delayed Posttest......................................................... 192
Table 32. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [-I] Group by Time: Immediate versus Delayed Posttest. ................................................. 192
Table 33. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [+I] Group by Time: Pretest versus Immediate Posttest...................................................... 192
Table 34. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [+I] Group by Time: Pretest versus Delayed Posttest......................................................... 193
Table 35. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [+I] Group by Time: Immediate versus Delayed Posttest. ................................................. 193
Table 36. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [++I] Group by Time: Pretest versus Immediate Posttest..................................................... 193
Table 37. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [++I] Group by Time: Pretest versus Delayed Posttest. ......................................................... 193
Table 38. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [++I] Group by Time: Immediate versus Delayed Posttest. ................................................. 194
Table 39. Means, SD, and SE for Meaning Recognition of Lexical Items Overall by Type of Glossing and Time ................................................................. 195

Table 40. ANOVA on Grammar Recognition by Time, Type of Item, and Group ....... 197

Table 41. Means, SD, and SE for Grammar Recognition by Type of Glossing, Type of Item and Time ........................................................................................................... 197

Table 42. One-Way ANOVA on Grammar Recognition within the [-I] Group by Time:
Pretest versus Immediate Posttest ................................................................. 198

Table 43. One-Way ANOVA on Grammar Recognition within the [-I] Group by Time:
Pretest versus Delayed Posttest ...................................................................... 198

Table 44. One-Way ANOVA on Grammar Recognition within the [-I] Group by Time:
Immediate versus Delayed Posttest ............................................................... 198

Table 45. One-Way ANOVA on Grammar Recognition within the [+I] Group by Time:
Pretest versus Immediate Posttest ................................................................. 199

Table 46. One-Way ANOVA on Grammar Recognition within the [+I] Group by Time:
Pretest versus Delayed Posttest ...................................................................... 199

Table 47. One-Way ANOVA on Grammar Recognition within the [+I] Group by Time:
Immediate versus Delayed Posttest ............................................................... 199

Table 48. One-Way ANOVA on Grammar Recognition within the [++I] Group by Time:
Pretest versus Immediate Posttest ................................................................. 199

Table 49. One-Way ANOVA on Grammar Recognition within the [++I] Group by Time:
Pretest versus Delayed Posttest ...................................................................... 200
Table 50. One-Way ANOVA on Grammar Recognition within the [++I] Group by Time:
Immediate versus Delayed Posttest. ................................................................. 200

Table 51. ANOVA on Meaning Production of Target Lexical Items by Time and Group
.................................................................................................................................. 202

Table 52. Means, SD, and SE for Meaning Production of Target Lexical Items by Type of
Glossing and Time ...................................................................................................... 203

Table 53. One-Way ANOVA on Meaning Production of Target Lexical Items within the
[-I] Group by Time: Pretest versus Immediate Posttest........................................ 203

Table 54. One-Way ANOVA on Meaning Production of Target Lexical Items within the
[-I] Group by Time: Pretest versus Delayed Posttest............................................. 204

Table 55. One-Way ANOVA on Meaning Production of Target Lexical Items within the
[-I] Group by Time: Immediate versus Delayed Posttest. ...................................... 204

Table 56. One-Way ANOVA on Meaning Production of Target Lexical Items within the
[+I] Group by Time: Pretest versus Immediate Posttest........................................ 204

Table 57. One-Way ANOVA on Meaning Production of Target Lexical Items within the
[+I] Group by Time: Pretest versus Delayed Posttest............................................. 204

Table 58. One-Way ANOVA on Meaning Production of Target Lexical Items within the
[+I] Group by Time: Immediate versus Delayed Posttest. ...................................... 205

Table 59. One-Way ANOVA on Meaning Production of Target Lexical Items within the
[++] Group by Time: Pretest versus Immediate Posttest....................................... 205

Table 60. One-Way ANOVA on Meaning Production of Target Lexical Items within the
[++] Group by Time: Pretest versus Delayed Posttest. ............................................ 205
Table 61. One-Way ANOVA on Meaning Production of Target Lexical Items within the 
[++I] Group by Time: Immediate versus Delayed Posttest. ................................. 205

Table 62. Means, SD, and SE for Meaning Production of Lexical Items Overall by Type 
of Glossing and Time............................................................................................. 207

Table 63. ANOVA on Grammar Production by Time, Type of Item and Group......... 208

Table 64. Means, SD, and SE for Grammar Production by Type of Glossing, Type of 
Item and Time........................................................................................................... 209

Table 65. One-Way ANOVA on Grammar Production within the [-I] Group by Time: 
Pretest versus Immediate Posttest......................................................................... 210

Table 66. One-Way ANOVA on Grammar Production within the [-I] Group by Time: 
Pretest versus Delayed Posttest ............................................................................ 210

Table 67. One-Way ANOVA on Grammar Production within the [-I] Group by Time: 
Immediate versus Delayed Posttest ...................................................................... 210

Table 68. One-Way ANOVA on Grammar Production within the [+I] Group by Time: 
Pretest versus Immediate Posttest ......................................................................... 210

Table 69. One-Way ANOVA on Grammar Production within the [+I] Group by Time: 
Pretest versus Delayed Posttest ............................................................................ 211

Table 70. One-Way ANOVA on Grammar Production within the [+I] Group by Time: 
Immediate versus Delayed Posttest ...................................................................... 211

Table 71. One-Way ANOVA on Grammar Production within the [++I] Group by Time: 
Pretest versus Immediate Posttest ......................................................................... 211
Table 72. One-Way ANOVA on Grammar Production within the [++] Group by Time: Pretest versus Delayed Posttest. .............................................................................. 211

Table 73. One-Way ANOVA on Grammar Production within the [++] Group by Time: Immediate versus Delayed Posttest. ................................................................. 212

Table 74. ANOVA for Text Comprehension by Type of Idea and Group on the Lexical Item Text................................................................................................................. 214

Table 75. Means, SD, and SE for Text Comprehension on the Lexical Item Text by Group and Type of Idea.......................................................................................... 214

Table 76. ANOVA for Text Comprehension by Type of Idea and Group on the Grammatical Item Text........................................................................................... 215

Table 77. Means, SD, and SE for Text Comprehension on the Grammatical Item Text by Type of Glossing.......................................................................................... 216

Table 78. Means, SD, and SE for Reported Memory Experiences by Type of Linguistic Item ......................................................................................................................... 217

Table 79. Means, SD, and SE for Reported Awareness by Type of Linguistic Item ..... 218

Table 80. Reported Awareness of Lexical Items by Reported Level of Awareness of Grammatical Items................................................................................................. 219

Table 81. ANOVA on Recognition by Type of Linguistic Item and Time ............... 221

Table 82. Means, SD, and SE for Recognition by Type of Linguistic Item and Time... 221

Table 83. One-Way ANOVA on Recognition of Lexical Items by Time: Pretest versus Immediate Posttest. .............................................................................................................. 222
Table 84. One-Way ANOVA on Recognition of Lexical Items by Time: Pretest versus Delayed Posttest ................................................................. 222

Table 85. One-Way ANOVA on Recognition of Lexical Items by Time: Immediate versus Delayed Posttest ......................................................... 222

Table 86. One-Way ANOVA on Recognition of Grammatical Items by Time: Pretest versus Immediate Posttest ............................................. 223

Table 87. One-Way ANOVA on Recognition of Grammatical Items by Time: Pretest versus Delayed Posttest ..................................................... 223

Table 88. One-Way ANOVA on Recognition of Grammatical Items by Time: Immediate versus Delayed Posttest ........................................... 223

Table 89. ANOVA on Production by Type of Linguistic Item and Time ............. 224

Table 90. Means, SD, and SE for Production by Type of Linguistic Item and Time ... 225

Table 91. One-Way ANOVA on Meaning Production of Lexical Items by Time: Pretest versus Immediate Posttest ......................................... 225

Table 92. One-Way ANOVA on Meaning Production of Lexical Items by Time: Pretest versus Delayed Posttest .............................................. 226

Table 93. One-Way ANOVA on Meaning Production of Lexical Items by Time: Immediate versus Delayed Posttest ....................................... 226

Table 94. One-Way ANOVA on Production of Grammatical Items by Time: Pretest versus Immediate Posttest .............................................. 226

Table 95. One-Way ANOVA on Production of Grammatical Items by Time: Pretest versus Delayed Posttest ................................................... 226
Table 96. One-Way ANOVA on Production of Grammatical Items by Time: Immediate versus Delayed Posttest.......................................................... 227
INTRODUCTION

Statement of the Problem

Models and theories in the second language acquisition (SLA) field present different views on the role of attention and awareness in adult second/foreign language (L2) development. While attention to the L2 input is deemed necessary for L2 learning by most researchers, the role of awareness seems to be more controversial. Several attentional models (Schmidt, 1990, 1993, 1994a, 1994b, 1995, 2001; Robinson, 1995b, 2003) posit that awareness at a low level is necessary and sufficient for intake, while awareness at a high level facilitates deeper learning. Moreover, from different theoretical perspectives, it has been suggested that awareness allows for encoding into long-term episodic memory (e.g., Robinson, 1995b; N. Ellis, 2001).

In the cognitive psychology field, Tulving (1983, 1985) was the first to introduce the notion of episodic memory. The researcher suggested that recognition memory retrieval is accompanied by different subjective experiences or states of awareness: remembering and knowing. Remembering involves recognition of an item with recollection of contextual details, and reflects episodic memory, while knowing involves recognition of an item without recollection of the context, and reflects semantic memory. Since Tulving’s proposal, a large number of studies in cognitive psychology and neuroscience have found evidence for the functional distinction of episodic and semantic memory, and their distinct neural correlates. Studies have shown that several variables, such as levels of processing, have a differential impact on experiences of remembering
and knowing. In particular, studies within the levels of processing framework have found
evidence to support a strong relationship between depth of processing and awareness, on
the one hand, and remembering of L1 linguistic items, on the other (e.g., Gardiner, 1988;
Rajaram, 1993; Java, 1994; Gardiner, Gawlik & Richardson-Klavehn, 1994; Richardson-
Klavehn & Gardiner, 1995; Gardiner, Java & Richardson-Klavehn, 1996; Karayianni &
Gardiner, 2003). In the SLA field, however, the relationship between awareness and
experiences of recognition memory, as well as the relationship between memory
experiences and L2 learning, has not been empirically investigated.

To measure participants’ memory experiences, Tulving (1985) proposed the
remember-know task, a procedure that has been widely employed in cognitive
psychology studies. To date, only three unpublished studies in SLA (Peckham, 2000;
Martínez-Fernández, 2009a, 2009b) have used this method. However, none of these
studies has examined the effects of levels of awareness, and findings regarding the
relationship between memory experiences and learning are very tentative.

Rather than delving into the relationship between awareness and memory, SLA
researchers have been more concerned with the roles of attention and awareness in
external conditions of exposure to the input. Since attentional models proposed a crucial
role for attention and awareness in L2 learning, researchers have extensively investigated
different pedagogical techniques, such as textual enhancement and glossing, to increase
the saliency of the linguistic items embedded in the input, and drive learners’ attention to
them. Studies on textual enhancement (i.e., typographical enhancement) have mainly
investigated effects on grammar learning. In contrast, studies on glossing (i.e., providing
a short definition of an obscure word at the bottom or at the margins of a text) have mainly examined effects on reading comprehension and vocabulary learning.

Overall, findings of glossing studies show that glossing is an effective technique to aid reading comprehension. Although there is some evidence to support a beneficial effect on vocabulary learning too, studies show that the vocabulary gains are generally very low (e.g., Bowles, 2004, Hulstijn, Hollander & Greidanus, 1996). Recently, researchers have started to investigate different types of glossing, such as glosses embedded in multiple-choice and fill-in-the-blank tasks, which are assumed to lead to deeper processing and higher vocabulary learning without interrupting the reading process (e.g., Watanabe, 1997; Hulstijn & Laufer, 2001; Rott, 2005; Martínez-Fernández, 2008a). Findings regarding the effectiveness of these lexical procedures, however, are not conclusive for either vocabulary learning or text comprehension. Moreover, although studies on glossing are premised on learners’ attention to and noticing, that is, being aware of the glosses, very few studies have attempted to measure attention, processing and/or awareness. Therefore, future research needs to establish whether different glossing conditions may impact learners’ cognitive processes, vocabulary development and text comprehension.

Finally, the use of glossing for grammatical items is recently gaining interest in the literature (e.g., Nagata, 1999; Guidi, 2009). In contrast to textual enhancement, which only provides enhancement of the forms in a text, glosses provide the meaning of forms as well as enhancement of the forms within the text. Thus, glossing might strengthen the form-meaning relationship of grammatical items, and have a greater impact than textual
enhancement alone on learners’ noticing and intake. To date, very few studies in the
glossing strand have examined the effects of type of linguistic item, which opens a door
for future research.

The present study aimed to investigate the effects of levels of awareness on
experiences of recognition memory and L2 development, as well as the effects of type of
glossing and type of linguistic item (i.e., vocabulary versus grammar) on awareness,
memory experiences, L2 development, and text comprehension.

**Definition of terms**

*Attention:* Following Robinson (1995b), the concept of attention can be used to
describe (a) “the process involved in “selecting” the information to be processed and
stored in memory”; (b) “the capacity for processing information”; and (c) the mental
“effort” involved in processing information” (pp. 287-288). In the present study, attention
is considered as a cognitive mechanism that learners use to select information, and which
is necessary to further process the information.

*Awareness:* The concept of awareness is used in this study from two different
perspectives. In the SLA field, awareness has been defined as “a particular state of mind
in which an individual has undergone a specific subjective experience of some cognitive
content or external stimulus” (Tomlin & Villa, 1994, p. 193). Within this framework,
awareness takes place at the stage where learners receive, process, and encode incoming
information. However, from the perspective of research studies on recognition memory,
the term *states of awareness* refer to the subjective experiences that accompany
individuals’ memory retrieval (Tulving, 1983, 1985). In other words, when an individual recognizes certain information he or she is aware of having encountered that information earlier, and his/her state of awareness will be different depending on whether it involves recollection of contextual details or not. In this sense, awareness (i.e., states of awareness or awareness of memory) does not take place at the encoding stage but at the retrieval stage, where individuals retrieve the information they have previously processed and encoded. It is important to note the difference between the two notions associated with the term of awareness since they constitute different variables in the present study. For the sake of clarity, states of awareness are typically referred to as experiences of memory in this study (except when the researcher’s words are intentionally used) and will refer to the retrieval or nonconcurrent stage.

Depth of processing: The concept of depth of processing was proposed by Craik and Lockhart (1972) in the cognitive psychology field to refer to conceptual or semantic processing (i.e., deep processing) versus perceptual processing (i.e., shallow processing). This concept has been refined during the last decades, and it is currently related to other notions such as amount of attention, degree of elaboration or level of awareness during processing rather than to semantic processing. The notion of depth of processing has been adopted in the SLA field to refer to amount of attention, mental effort, and/or noticing induced by different types of tasks (cf. Robinson, 2003; Hulstijn, 2001; Laufer & Hulstijn, 2001; Leow, Hsieh, & Moreno, 2008).

Gloss: A gloss is traditionally defined as a definition, synonym or translation that is provided for a particular obscure word at the bottom or margin of a text in order to
improve text comprehension. However, this notion has been extended to include different types of annotations (e.g., audio, image, video) provided in order to aid text comprehension and/or promote learning.

**Episodic/Semantic memory:** Episodic memory refers to the memory of personally experienced events (including times, places, associated emotions, and other contextual knowledge) that can be explicitly stated. Semantic memory refers to the memory of facts, meanings, and general knowledge about the world that is unrelated to specific experiences, and can be explicitly stated. The notions of episodic and semantic memory should be distinguished from the notions of episodic and semantic memory systems. Some researchers support the idea that episodic memory and semantic memory are stored in different memory systems in the brain (cf. Baddeley, Conway, & Aggleton, 2002; Gardiner, 2008). Other researchers, however, consider episodic and semantic memory as different processes unrelated to memory systems (cf. Yonelinas, 2002). In this study, no assumption is made regarding the nature of episodic and semantic memory as processes or memory systems.

**Incidental/Non-incidental conditions of exposure:** External conditions by which learners are exposed to specific linguistic items in a communicative context where the main goal is input comprehension rather than grammar or vocabulary learning (+incidental condition), or in a less communicative context that requires learners to focus on the linguistic items to some extent (-incidental condition). External conditions of exposure must be distinguished from the internal processes that learners may engage in
(i.e., learners may or may not pay attention to target linguistic items and become aware of them regardless of their condition of exposure).

**Input enhancement:** A number of techniques used to make selected features of the input more salient in such a way as to promote noticing and facilitate incidental learning of targeted forms when the main goal is to comprehend the input (Sharwood Smith, 1991).

**Input:** In the context of language learning, input refers to the language that the learner is exposed to and from which he or she can learn.

**Intake:** The part of the input that has been attended to during input processing. Following Leow (1993), “intake represents stored linguistic data that may be used for immediate recognition and does not necessarily imply language acquisition” (p. 334).

**Long-term memory:** Long-term memory refers to the memory system or processes responsible for encoding, deletion, consolidation, and retrieval (i.e., recall and recognition) of information, events, and skills. In the present study, long-term memory is distinguished from the notions of long-term effects and performance on delayed posttests typically used in the SLA field. Following cognitive psychology research, long-term memory can be measured a few minutes after exposure or decades later.

**Noticing:** Based on Schmidt (1994a), noticing is defined as the conscious registration of the occurrence of a stimulus. Within this framework, noticing involves attention and a low level of awareness, results in intake, and is deemed crucial and necessary for learning to take place. Additionally, noticing is considered to involve rehearsal in working memory prior to encoding in long-term memory (Robinson, 1995b).
**Remembering/Knowing:** Based on Tulving (1983, 1985) remembering and knowing are different subjective states of awareness that accompany memory retrieval. Remembering is an experience characterized by the specific recollection of contextual details involving one’s self at a particular time, in a particular place. In contrast, the experience of knowing is characterized by a more abstract awareness of familiarity, without any re-living of past events in which one was personally involved. The experiences of remembering and knowing are assumed to reflect episodic and semantic memory, respectively. The remember-know distinction refers also to one of the techniques used to investigate these memory experiences.

**Textual enhancement:** A technique of input modification consisting on typographical or visual enhancement of the input in written texts, such as color-coding, boldfacing, highlighting, or manipulations of the font and the character size.

**Type of glossing:** Glossing is a technique used to aid learners’ text comprehension and L2 development while reading. Researchers have investigated different types of glossing in order to facilitate L2 development, including the following: (a) traditional glossing, which provides a definition, synonym or translation for a targeted word (e.g., Bowles, 2004); (b) multiple-choice glosses, which display several translations from which the reader chooses (e.g., Hulstijn, 1992); (c) fill-in tasks, which display several glossed words from which the reader chooses the appropriate word to fill in the blanks of a text (e.g., Hustijn & Laufer, 2001); and (d) gloss-retrieval, which involves glossing of the targeted word during the first encounter and answering a question about its meaning in a subsequent encounter (e.g., Rott, 2007). In the present study, two types of glossing
are investigated: traditional glosses and glosses embedded in a fill-in-the-blank task. Both
types of glossing in this study involve textual enhancement (i.e., bolding) within the text
and in the glosses provided in the margin, as is often done in pedagogical textbooks.
CHAPTER 1: REVIEW OF RELATED LITERATURE

Theoretical Foundation

In the last decades, several theoretical models have been proposed in the cognitive psychology field to explain long-term recognition memory experiences and their relationship with awareness: Tulving’s model of episodic and semantic memory systems (Tulving, 1983, 1985); dual-process models, such as Yonelina and Jacoby’s model of recollection and familiarity (Jacoby, 1991; Yonelinas & Jacoby, 1995), and Rajaram’s model of distinctiveness and fluency (Rajaram, 1993, 1996, 1998); single-process models, such as Donaldson’s signal detection model (Donaldson, 1996); and other proposals that have been recently developed (e.g., Rotello, MacMillan, & Reeder, 2004). Although experiences of recognition memory have not received much attention in the SLA field, the following SLA models and proposals recognize the importance of long-term memory in L2 development: Schmidt’s noticing hypothesis (Schmidt, 1990, 1993, 1994a, 1994b, 1995, 2001); Robinson’s model of the relationship between attention and memory (Robinson, 1995b, 2003); N. Ellis’ model of memory for language (N. Ellis, 2001); and Ullman’s model of declarative/procedural memory (Ullman, 2001). Since models in both the cognitive psychology and the SLA fields are relevant to the theoretical underpinnings of the present research, the major ones are described in the sections below.
Models of Recognition Memory Experiences in Cognitive Psychology: Remembering and Knowing

Over the last 30 years, different theoretical models employing a variety of methods have emerged to explain the difference between two experiences of memory: remembering, which reflects recollection of contextual details, and knowing, which does not involve such recollection. These models present different theoretical perspectives regarding the underlying nature of memory experiences, their characteristics, and the relationship between them. Dual-system (e.g., Tulving, 1983, 1985) and dual-process theories (e.g., Jacoby, 1991; Yonelinas & Jacoby, 1995; Rajaram, 1993, 1996, 1998) consider these experiences as qualitatively different. In contrast, single-process theories (e.g., Donaldson, 1996) claim that memory experiences merely reflect different strength of the same process. The next sections present a review of the major models that have been proposed in the literature.

Tulving’s model of episodic and semantic memory systems

Tulving (1983, 1985) was the first to suggest that memory retrieval could be accompanied by different subjective states of awareness (i.e., experiences of memory): remembering and knowing. Remembering is an experience characterized by the specific recollection of contextual details involving one’s self at a particular time, in a particular place. In contrast, the experience of knowing is characterized by a more abstract awareness of familiarity, without any re-living of past events in which one was personally involved. Tulving (1983) considered the experience of remembering as an expression of
‘autonoetic consciousness’ (i.e., self-knowing consciousness), and the experience of knowing as an expression of ‘noetic consciousness’ (i.e., knowing consciousness). In addition, he argued that these types of consciousness were properties of two functionally distinct memory systems, episodic and semantic memory systems, respectively. The episodic memory system stores personally experienced events and their temporal relations to each other, and the semantic memory system stores general knowledge about the world. In this view, while autonoetic and noetic consciousness reflect episodic and semantic memory respectively, and both involve awareness of memory, a third type of consciousness, ‘anoetic consciousness’ (i.e., not knowing consciousness), reflects procedural memory, which does not involve awareness of memory, and which relates to the procedural memory system.

In Tulving’s (1995) model, encoding of events is serial (i.e., all events are first encoded into the semantic system, and then into the episodic system), storage may be parallel (i.e., events can be stored in both systems at the same time), and retrieval is independent (i.e., events can be retrieved from either system). When events are only minimally registered, they may only be stored in the semantic system. However, the more attention and conscious control at encoding, the more chances are that the event will also be encoded and stored in the episodic system. Tulving (1993) proposed the coordination hypothesis, which posits a relationship between awareness at encoding and awareness at retrieval. According to this hypothesis, however much effort goes into the retrieval attempt, the resulting state of awareness (i.e., remembering or knowing) cannot exceed the level of awareness achieved at encoding (i.e., encoding into both memory systems or
encoding into the semantic system only). In other words, events that are only encoded into the semantic system cannot be retrieved with an experience of remembering.

Tulving’s model has several implications for learning. Both the episodic and semantic systems are expected to support learning of new information but they play different roles. While recall performance relies more on the episodic system, recognition relies on both systems. On the one hand, the semantic system is considered to be slower to learn new information, but can support novel learning even without the episodic system. On the other hand, episodic memory involves higher attention and awareness at encoding, and is more efficient to learn new information; however, it cannot support novel learning without the semantic system. Tulving’s model assumes that attention and awareness are involved in both semantic and episodic memory but assigns a higher impact of these constructs on episodic memory (i.e., the more attention and awareness during encoding, the higher episodic memory).

In order to measure the experiences of remembering and knowing, Tulving (1985) developed a specific judgment task. In this study, participants were presented with a list of words, and were asked to perform a particular encoding task; then they were given a longer list containing the original words as well as new words, and were asked to (a) recognize items as old or new, and (b) judge whether they recognized old items on the basis of remembering or knowing (by selecting a ‘remember’ or ‘know’ response). To obtain estimates of memory experiences, the number of remember responses given to new items (i.e., ‘false alarms’) were subtracted from the number of remember responses given to old items (i.e., ‘hits’), and the same procedure was followed for know responses.
This method has been adopted in many empirical studies, giving rise to the so-called remember-know paradigm. Although Tulving (1985) used the remember-know procedure with free recall, cued recall, and recognition tasks under incidental and intentional conditions, further investigations have mainly focused on recognition tasks.

Researchers have progressively refined the method to measure recognition memory experiences. In early studies, the proportions of remember and know responses were used as estimates of the processes underlying remembering and knowing, respectively. However, because participants were allowed to give only one response, this procedure assumed that the underlying processes were exclusive. Moreover, participants could give a know response for an item that was remembered because it was also familiar. Therefore, the instructions for this task were modified so that participants would select know responses for items that were “familiar and not remembered” (Gardiner & Richardson-Klavehn, 2000).

Other modifications of the original version of the procedure include breaking down ‘know’ responses into additional response categories. Gardiner, Java, and Richardson-Klavehn (1996) allowed guesses to be reported as ‘guess’ responses, a practice that has been adopted later in a large number of studies. The addition of guess responses can remove a potential confounding of recognition decisions that are more strategically based with responses that are based on memory experiences. Finally, Conway, Gardiner, Perfect, Anderson, and Cohen (1997) distinguished ‘know’ responses, which would reflect a strong feeling of familiarity, from ‘familiarity’ responses, which would reflect a weak feeling of familiarity. This version of the procedure, however, has
not been generalized to other studies. As Gardiner and Richardson-Klavehn (2000) indicate, the different measures used so far pose the question whether there are two experiences of recognition memory - remembering and knowing - or more. Furthermore, subjective reports may be a problematic measure because participants may not be able to accurately distinguish between experiences of remembering and knowing. For this reason, in most experiments participants are given written instructions with detailed descriptions of the memory experiences, which are followed by a discussion to ensure comprehension. Additionally, in order to check whether participants follow the instructions, they are asked to provide descriptions of the reasons underlying a random selection of responses after taking the test. Yet, as Gardiner (2008) has recently indicated, neither the use of posttest checks of the validity of the responses nor the use of guess responses has become universal practice, which increases the risk of obtaining potentially misleading results. More importantly, research studies on memory experiences seem to assume that offline data such as posttest checks have a direct relationship with online processes such as awareness of memory (cf. Leow, Johnson, & Zárate-Sández, forthcoming for a discussion on online and offline measures).

In spite of these methodological issues, the procedure developed by Tulving (1985) has led to systematic results, which suggests that participants tend to accurately and consistently distinguish between remembering and knowing experiences. A number of empirical studies have found support for Tulving’s model (e.g., Gardiner, 1988; Gardiner & Parkin, 1990; Gregg & Gardiner, 1994; Gardiner, Java, & Richardson-Klavehn, 1996; Gardiner, Ramponi, & Richardson-Klavehn, 1999; Konstantinou &
Gardiner, 2005; Gardiner, Gregg, & Karayianni, 2006; Gardiner, Brand, Vargha-Khadem, Baddeley, & Mishkin, 2006). In particular, findings indicate that remembering and knowing are differently affected by several independent variables, such as levels of processing, divided attention, and retention intervals (e.g., Gardiner, 1988; Gardiner & Parkin, 1990; Rajaram, 1993; Konstantinou & Gardiner, 2005). This evidence supports the idea that the processes underlying experiences of remembering and knowing are qualitatively different. Furthermore, studies have shown that these experiences are differently affected by the use of alcohol and drugs (e.g., Curran & Hildebrant, 1997) and that vary systematically with respect to different subjects: old versus young adults (e.g., Perfect, Williams & Anderton-Brown, 1995), and healthy subjects versus amnesic, epileptic, schizophrenic, autistic, and Alzheimer’s patients (cf. Gardiner, 2008 for a review). Interestingly, these dissociations in memory experiences cannot be inferred from traditional measures of recognition memory performance.

Finally, some evidence from lesion studies as well as from event-related potential (ERPs) and functional imaging (fMRI) studies suggests that experiences of remembering and knowing have partially distinct neural correlates (Knowlton & Squire, 1995). This result supports Tulving’s prediction that the relationship between memory systems is one of inclusiveness or redundancy (i.e., episodic memory is qualitatively different from semantic memory but cannot occur without semantic memory). On the one hand, the literature shows that both experiences are dependent on the medial-temporal lobe, which is essential for declarative memory. For example, some studies have shown that amnesic patients, who have medial-temporal lobe damage, cannot experience either remembering
or knowing (Knowlton & Squire, 1995). On the other hand, there is increasing evidence that these experiences are associated with different patterns of neural activity within the medial temporal lobe. Although the functional specialization of the medial temporal lobe is still controversial, some studies suggest that the hippocampus selectively supports remembering but not knowing, while regions in the surrounding of the hippocampus, such as the perirhinal cortex, are crucial for the experience of knowing (cf. Aggleton & Brown, 2006, and Eichenbaum, Yonelinas, & Ranganath, 2007 for a review).

**Dual-process models**

Some models of recognition memory propose that memory experiences reflect two kinds of processes rather than two memory systems. Within a dual-process framework, there have been two major proposals. First, Jacoby (1991), Yonelinas and Jacoby (1995), and Jacoby, Yonelinas, and Jennings (1997) elaborated a model to account for processes of ‘recollection’ and ‘familiarity’, assumed to underlie the experiences of remembering and knowing, respectively. Second, Rajaram (1993, 1996, 1998) proposed a different dual-process model that related experiences of remembering and knowing to ‘distinctiveness’ and ‘fluency’ of processing, respectively.

The recollection and familiarity model differs from Tulving’s model minimally in the following aspects: (a) argues that both processes are independent rather than redundant, (b) considers recollection as a slow, effortful process that depends on conscious control at retrieval, while familiarity is seen as a fast, automatic, and unconscious process, and (c) assumes that recollection supports high confidence
recognition decisions, while familiarity supports a wide range of recognition confidence responses (cf., Yonelinas, 2002). Measures used to test this model include the remember-know procedure,¹ but also other measures such as the response-deadline procedure, the process-dissociation procedure (PDP), and the receiver operating characteristics (ROC curves). The response-deadline procedure is a task-dissociation method (i.e., a task or test condition is assumed to isolate one of the two processes) in which participants are given different amount of time to make recognition judgments; fast responses are assumed to reflect familiarity, while slow responses would reflect recollection. This type of method can lead to ambiguous results since estimates of the processes being measured are imprecise. In contrast, process-estimation methods (e.g., the remember-know procedure, the process-dissociation procedure, and ROC curves) provide quantitative estimates of the contribution of each process to overall recognition performance.

The process-dissociation procedure was developed by Jacoby (1991) to account for the independence of processes. In this procedure, participants study a visually presented list of items, then a separate list of heard items, and then they complete two different tests: (a) the inclusion test, in which they must respond ‘yes’ to items previously encountered (in any of the lists), and (b) the exclusion test, in which they respond ‘yes’ only to items from the heard list. Recollection is measured as the ability to remember where or when an item was earlier studied (i.e., performance on the exclusion test). Both recollection and familiarity processes are assumed to contribute to performance on the

¹ A new version of the paradigm that allows independence between remember and know responses was developed by Yonelinas and Jacoby (1995).
inclusion test, although it is not clear to what extent each process contributes to overall recognition.

In the ROC method, participants are typically asked to make old/new judgments, give remember-know responses, and rate the confidence of their recognition responses, so that the proportion of hits (i.e., old items correctly identified as old) to false alarms (i.e., new items incorrectly identified as old) under varying confidence levels can be calculated. A ROC is the plot of the hit rate (the rate at which old items are correctly identified as old) versus the false alarm rate (the rate at which new items are incorrectly identified as old) for decisions made with different levels of confidence.

Findings from lesion studies and ERP and fMRI studies have also contributed to develop this model. Recently, Bowles, Crupi, Mirsattari, Pigott, Parrent, Pruessner, Yonelinas and Kohler (2007) conducted four experiments in which recollection and familiarity were measured by three methods (remember-know procedure, ROC, and response-deadline procedure) in an individual who underwent surgical resection of a large portion of perirhinal cortex that spared the hippocampus for treatment of epilepsy. Results showed impaired familiarity with preserved recollection, thus supporting the idea that the hippocampus is involved in the process of recollection, while the perirhinal cortex, in the surrounding area of the hippocampus, is crucial for familiarity.

In spite of the differences with respect to the memory systems model, the recollection and familiarity model also considers a qualitative distinction between memory experiences, supports the effect of levels of processing on memory experiences (although in this account memory experiences may also depend on awareness at retrieval.
and not only on awareness at encoding), and assesses a stronger impact of recollection on learning.

Finally, the distinctiveness and fluency model can complement both the memory systems and the recollection and familiarity models. According to Rajaram (1996, 1998) remembering benefits from distinctiveness of processing, which involves greater attention and more elaborative processing, and knowing benefits from fluency of processing, which implies less attention and minimal processing. While distinctiveness may increase encoding into episodic memory or enhance a recollection process, fluency may lead to encoding into semantic memory (without encoding into episodic memory) or enhance a familiarity process.

**Single-process models**

Donaldson (1996), Hirshman and Master (1997), and Inoe and Belleza (1998) proposed single-process accounts to explain the differences between remembering and knowing, based on a signal detection model. The signal detection model postulates that the difference between remembering and knowing is not qualitative, but rather quantitative, since these experiences are assumed to correspond to different regions on a single memory strength continuum. According to this model, remembering reflects a stronger trace and is associated with high confidence in the recognition judgments, while knowing reflects a weaker trace and is associated with lower levels of confidence. Therefore, responses on the remember-know task are considered to be equivalent to confidence judgments. Moreover, researchers that support this model suggest that
remember and know responses might merely reflect decision processes in recognition memory: participants using more stringent criteria would select know responses more often, while participants using more lenient criteria would select remember responses, the strength of the trace being the same in both cases. The signal detection model has been strongly criticized from different perspectives (e.g., Gardiner & Gregg, 1997; Gardiner, Ramponi & Richardson-Klavehn, 2002; MacMillan, Rotello & Verde, 2005). Most critiques indicate that this model cannot explain the following results: (a) functional independence of remembering and knowing, and why different independent variables produce specific dissociations and associations between these experiences, (b) empirical findings showing no correlation between these experiences and confidence judgments, and (c) evidence based on neural patterns associated with each process.

However, single-process accounts have recently gained interest (e.g., Dunn, 2004; Wixted & Strecht, 2004; Squire, Wixted, & Clark, 2007). After examining findings from lesion and neuroimaging studies within a single-process framework, Squire et al. (2007) have suggested that both the hippocampus and the perirhinal cortex equally contribute to recollection and familiarity, which they interpret as strong and weak memory, respectively. According to them, “strong memories are associated with increased hippocampal activity, regardless of whether they reflect strong familiarity, strong recollection or a combination of the two. Further, weak memories are often not associated with detectably increased activity in the hippocampus, regardless of whether such memories reflect weak familiarity, weak recollection, or a combination of the two” (p. 878). In their view, fMRI techniques are relatively insensitive to detect weak memory...
(i.e., familiarity or knowing in two-process theories) in the hippocampus, and strong memory (i.e., recollection or remembering in two-process theories) in the perirhinal cortex. It is not clear, however, how this interpretation could account for recent findings reported by Bowles et al. (2007), which show that the perirhinal cortex, whether usually involved in recollection or not, is clearly not necessary for this process to take place, while the hippocampus seems to be crucial (as it is generally accepted). In any case, both dual-process and single-process models are currently being tested, and further research should shed more light on this issue.

Recent developments

As can be inferred from the previous review, after 30 years of research there is still much controversy regarding the nature of and the relationship between the experiences of remembering and knowing, and between the processes of recollection and familiarity. The contribution of lesion and neuroimaging studies has considerably enriched the understanding of these processes, although further research is needed in order to consolidate findings. Most studies have found evidence for dual-system or dual-process models, but recently single-process theories seem to have gained interest in the literature. This process of constant review, evaluation and critique has made researchers revise original positions, refine methods, and propose new models.

Overall, new models tend to support dual-component accounts, while taking advantage of mathematical and computational models (e.g., Rotello, MacMillan, & Reeder, 2004; Reder, Nhoyvanisvong, Schunn, Ayers, Angstadt & Hirakiand, 2000).
Rotello, Macmillan, and Reeder (2004), for example, have critiqued the signal detection model as well as dual-component models, and have proposed a two-dimensional signal-detection model (the sum-difference theory of remembering and knowing model). This model assumes that remembering and knowing reflect the contributions of two distinct processes, but retains the advantages of the signal detection model (e.g., distinguishing accuracy from response bias and accounting for levels of confidence). In their study, participants rate their memories not only on an old-new confidence scale but also on a scale representing the degree to which their memory feels like remembering or knowing. ROC curves are generated by both ratings. The assumption is that memory strength varies not only on a global strength dimension, but also on a dimension that measures the specific strength of details associated with test items. In contrast with Yonelinas and Jacoby’s (1995) model, where only familiarity is assumed to reflect a variety of levels of confidence, in this model both recollection and familiarity may reflect different degrees of confidence, depending on the strength of details associated with items.

To date, however, there are no qualitative studies that investigate the nature and characteristics of experiences of remembering and knowing. Future studies will have to fill in this gap by using both quantitative and qualitative methods to investigate experiences of recognition memory. Although the present study is not concerned with these theoretical aspects of the research on memory experiences, it does acknowledge the importance of incorporating qualitative methods to measure memory experiences, such as participants’ concurrent verbalizations while performing a remember-know task.
Conclusion

In summary, since Tulving (1983) suggested the existence of two experiences or states of awareness associated with recognition memory –remembering and knowing–, a growing number of research studies has been conducted to investigate their nature, characteristics and implications. Researchers have proposed different theoretical models to explain the difference between remembering and knowing. Although there is no agreement on whether these experiences reflect two distinct processes or the same process, the idea that the processes underlying experiences of remembering and knowing are qualitatively different is well supported by the literature. In addition, as Gardiner (2008) concludes, it is important that models explain how different experiences of consciousness come to mind. The single-process accounts proposed so far only explain how decision processes may operate in responses regarding memory experiences, once these experiences have taken place. Instead, dual-component models provide a more psychologically meaningful account by relating memory experiences to different levels of awareness at encoding.

Finally, research within both single-process and dual-component frameworks have focused on theoretical issues, and very few studies have applied these models to areas such as learning (e.g., Conway et al., 1997). The field of second language acquisition could greatly benefit from research conducted in psychology over the last 30 years. Specifically, dual-component models of recognition memory proposed in cognitive psychology may shed light on the role of awareness in L2 development, a controversial issue that is currently being debated in the SLA field.
Models of Memory and/or Awareness in SLA

Models and hypotheses in the SLA field present different views on the role of awareness in L2 learning (e.g., Schmidt, 1990; Tomlin & Villa, 1994; Robinson, 1995b). Currently, most researchers recognize the importance of attention in L2 development, but there is no general agreement on the role of awareness. Awareness has been defined as “a particular state of mind in which an individual has undergone a specific subjective experience of some cognitive content or external stimulus” (Tomlin & Villa, 1994, p. 193). SLA researchers have related awareness to both short-term and long-term memory, although the latter has received little attention in the literature. The next sections present a review of models that focus on awareness and/or long-term memory in SLA within attentional, connectionist and neuroscientific frameworks. In particular, the following models are briefly summarized: Schmidt’s noticing hypothesis (Schmidt, 1990, 1993, 1994a, 1994b, 1995, 2001); Robinson’s model of the relationship between attention and memory (Robinson, 1995b, 2003); N. Ellis’ model of memory for language (N. Ellis, 2001); and Ullman’s model of declarative/procedural memory (Ullman, 2001).

**Schmidt’s Noticing Hypothesis**

Schmidt (1990, 1993, 1994a, 1994b, 1995, 2001) distinguished two levels of awareness: awareness at a low level, ‘noticing’, and awareness at a high level, ‘understanding.’ The researcher proposed the noticing hypothesis, which posits that attention is necessary for noticing, and that noticing is necessary and sufficient for intake, while understanding facilitates deeper learning. The crucial role of awareness is
controversial, since other researchers claim that it is possible to learn without awareness (e.g., Tomlin & Villa, 1994; Williams, 2004, 2005). According to Schmidt (2001), however, what is relevant may not be if awareness is necessary or not, but rather the fact that more awareness results in more learning.

In this account, awareness is operationalized as ability for verbal report during or after exposure to the input, although Schmidt (1990 and elsewhere) admits that failure to provide a verbal report does not necessarily reflect absence of noticing, since individuals may differ in their ability to verbalize contents of awareness. With respect to retroactive reports, recognition tests are considered a more accurate measure than recall tests because memory can affect to a greater extent the recall of forms previously noticed. Schmidt (2001) indicates that “failure to achieve above-chance performance in a forced-choice recognition test is a much better indication that the subjective threshold of perception has not been exceeded and noticing did not take place [compared to failure to recall on recall tests]” (p. 20). Retroactive recognition measures are also controversial because it may be possible to notice and immediately forget, thus failing to recognize on a subsequent test. However, it is important to note that the researcher suggests a strong relationship between noticing and recognition memory. A relationship between awareness and long-term memory becomes explicit in Schmidt’s (1994a) definition of noticing: “the registration of the occurrence of a stimulus event in conscious awareness and its subsequent storage in long-term memory” (p. 166). Although several studies have recently investigated the role of awareness in SLA (e.g., Alanen, 1995; Hama & Leow, 2010; Leow, 1998, 2001a; Robinson, 1995a, 1996, 1997a, 1997b; Williams, 2004, 2005) and some have found
support for a beneficial effect of higher levels of awareness (Leow 1997a, 2000, Rosa & Leow, 2004, Rosa & O’Neill, 1999), to date there has been almost no discussion of the relationship between awareness and long-term memory proposed in this framework.

Robinson’s Model of the Relationship Between Attention and Memory

Robinson (1995b, 2003) also supports the noticing hypothesis, claiming that no learning can take place without attention and some level of awareness. The researcher reviews different models of short-term and long-term memory in cognitive psychology to propose a model of attention and memory that is consistent with Schmidt’s hypothesis. According to him, noticing requires rehearsal in working memory prior to encoding in long-term memory: “It is possible to briefly notice and permanently or temporarily forget, and to notice and remember over time. More permanent encoding in long-term memory is a consequence of the level of activation of information, itself the result of rehearsal and elaboration” (Robinson, 1995b, p. 298). In this model, short-term/working memory is seen as a subset of long-term memory in a currently activated state, and the place where noticing takes place. In addition, rehearsal in short-term memory is assumed to depend on the processing demands of the tasks performed.

While SLA research is currently investigating the role of short-term/working memory (e.g., Robinson, 2002; Kormos & Sáfár, 2008; Medina, 2009), long-term memory has generally received little attention in the SLA field. In his review, Robinson (1995b, 2003) describes several models of long-term memory proposed in cognitive psychology, such as multiple-system models and processing models, among others.
Multiple-system models include (a) models of declarative memory (i.e., consciously accessed system) versus procedural memory (i.e., unconsciously accessed system), and (b) models of episodic memory (i.e., memory system for personal involvement) versus semantic memory (i.e., memory system for decontextualized information), both of them within the declarative system and outside the procedural memory system. This distinction between memory systems can be interpreted from a processing view as a functional distinction not necessarily related to different systems. Robinson (1995b, 2003) argues that processing views are more adequate because they avoid the confusion raised by multiple-system models. In any case, the model of long-term memory that has been more extensively investigated in SLA is the declarative/procedural model (e.g., Paradis, 1994; Ullman, 2001). Thus, further research needs to delve into the role of episodic and semantic memory in SLA, and investigate the relationship between memory experiences, awareness, and learning. Although Robinson (1995b) points out a relationship between episodic memory and noticing, as he states that “what is noticed may be subsequently encoded in long-term episodic memory” (p. 298), this relationship still needs to be empirically investigated in the area of second language learning.

N. Ellis’s Model of Memory for Language

Building upon theories of cognitive psychology, cognitive neuroscience, and computational models, N. Ellis (2005) proposes a connectionist model where attention, awareness and episodic memory are related. The researcher claims that the large majority of language learning takes place implicitly, resulting from a general associative process
where frequency and saliency of the input play a major role (N. Ellis 2001, 2005, 2007). However, in this view explicit learning is necessary for initial registration of novel linguistic constructions (i.e., form-meaning associations). According to N. Ellis’ (2001) model of memory for language, neural systems in the prefrontal cortex involved in working memory are required to temporarily perceive, register and notice new features or events; neural systems in the hippocampus bind these features into consciousness, and consolidate them into a long-term episodic memory; and finally these episodic representations are adopted by other brain regions in the neocortex, and integrated into the system by implicit learning during subsequent input processing.

For the contribution of conscious working memory in the formation of new long-term memories, N. Ellis (2005) reports evidence found in cognitive psychology and neuroimaging studies. On the one hand, cognitive psychology studies show that encoding of explicit memories is hindered by divided attention during learning, and enhanced by preparatory attention and voluntary orienting (e.g., Craik, Govoni, Naveh-Benjamin, & Anderson, 1996; Cowan, 1997). On the other hand, neuroimaging studies indicate that the amount of frontal activation at encoding (associated with deep elaborative encoding) predicts subsequent memory of verbal material (e.g., Wagner et al., 1998).

Consistent with these findings, and in line with research on episodic long-term memory, Baddeley (2000) developed a new version of his model of working memory,2 which fits N. Ellis’s (2001) proposal. Baddeley’s new model of working memory

---

2 Baddeley and Hitch (1974) proposed the three-component model of working memory. In this model, ‘working memory’ is defined as the limited capacity system that allows the temporary storage and manipulation of information necessary for complex tasks, such as comprehension, learning, and reasoning. Their model comprised three components: an attentional control system, the ‘central executive’, aided by two subsidiary slave systems, the ‘phonological loop’ and the ‘visual-spatial sketchpad’.
comprises four components: a supervisory attentional system, with two temporary slave subsystems for perceiving and representing both auditory and visual information (the ‘phonological loop’ and the ‘visual-spatial sketch pad’), and a temporary storage system, named the ‘episodic buffer’. The researcher indicates that the role of this new component, the episodic buffer, is binding information from the phonological loop and the visual-spatial sketch pad subsystems of working memory with information from episodic long-term memory. Therefore, this model explains how conscious working memory may contribute in the formation of new long-term memories.

Finally, it is important to note that N. Ellis’s (2001) model of memory for language predicts no distinction between L2 vocabulary and grammar learning. According to the researcher, lexical learning involves perception of frequent exemplars and sequences in the phonological loop, which allows chunking (i.e., development of permanent sets of associative connections) in long-term memory. In turn, long-term memory allows more ready perception of those exemplars and regular sequences in the input, which will be noticed and represented in working memory. The same process is applied to sequences of words, phrases, idioms, and the whole complex hierarchical system of language. Thus, regular morphology will be learned in the same way as irregular morphology. Learners do not apply rules, but rather they abstract the regularities of the language from sequences that are repeated in the input, and whose representations become connected in a neural network. In N. Ellis’s (2001) words, “syntax acquisition reduces to vocabulary acquisition” (p. 54).
In sum, N. Ellis (2001) relates findings and models in cognitive psychology and neuroscience to SLA, and explicitly recognizes the importance of episodic long-term memory in L2 vocabulary and grammar learning. However, as has been mentioned above, almost no empirical or neuroscientist studies have been conducted yet in the SLA field to support the role of episodic memory.

**Ullman’s Model of Declarative/Procedural Memory**

Within the field of neuroscience, Ullman (2001) proposed the model of declarative/procedural memory to account for both first and second language acquisition. In stark contrast with connectionist models, he argues that learning of lexicon and grammar in the first language are mediated by two different memory systems, namely, the declarative system and the procedural system, respectively. On the one hand, the declarative system is implicated in the learning, representation and use of knowledge about facts (i.e., semantic knowledge) and events (i.e., episodic knowledge), which is consciously accessed. This system is rooted in the medial temporal lobe regions (the hippocampus and related structures), and subserves an associative memory that facilitates lexical learning. On the other hand, the procedural system is implicated in the learning and control of motor and cognitive skills or habits, which cannot be consciously accessed, and is rooted in frontal/basal-ganglia structures.

According to Ullman (2001), procedural memory is more sensitive to age of exposure than declarative memory. For this reason, late language learners tend to rely on declarative memory to learn L2 grammar, and not on procedural memory. In this way,
regular morphologically complex forms may be memorized in the lexicon instead of being compositionally computed, just like irregular forms or other words. In addition, some grammatical rules may be learned in declarative memory. Therefore, dissociations between lexicon and grammar should be weaker in L2 than in L1. Yet, the model does not rule out the role of procedural memory in L2 learning, since practice with L2 may increase the degree of dependence on this system for grammatical computations.

Ullman’s model differs from others, such as those proposed by Paradis (1994) and N. Ellis (2001). Paradis’ model also emphasizes a shift from procedural memory in L1 to declarative memory in L2. However, it does not predict dissociations between lexicon and grammar in the L1 (i.e., the procedural system subserves L1 lexical learning). In N. Ellis’s (2001) model, L2 vocabulary and grammar learning take place in declarative memory too, but this model does not propose a shift from procedural to declarative memory, since L1 vocabulary and grammar learning also depends upon associative declarative memory. Although for different reasons, however, researchers seem to agree on the crucial role of declarative memory in L2 lexical and grammatical development. Therefore, future studies should delve into the role of this system, and investigate whether episodic and semantic memory systems, both within declarative memory, may affect L2 development in different ways.

Conclusion

Building upon cognitive psychology and neuroscience research, different theoretical models in SLA have suggested a relationship between awareness and episodic
memory (Schmidt, 1990 and elsewhere; Robinson, 1995b, 2003; N. Ellis, 2001), and have emphasized the role of declarative long-term memory in L2 development (Ullman, 2001). To determine the extent to which awareness may have an effect on memory and learning is crucial to understand second language learning from a cognitive perspective. Empirical studies in the cognitive psychology field have investigated the declarative memory system, specifically the role of episodic and semantic memory and experiences of remembering and knowing. However, this task still needs to be undertaken by SLA research. The next sections present a review of empirical studies that investigate the relationship between awareness and experiences of remembering and knowing.
Empirical Studies on the Relationship Between Memory Experiences and Awareness

Empirical Studies in Cognitive Psychology

Cognitive psychology research has examined the effects of levels of processing on experiences of remembering and knowing in a large number of empirical studies conducted over the last decades (e.g., Gardiner, 1988; Gardiner, Java, & Richardson-Klavehn, 1996; Gardiner, Brand, Vargha-Khadem, Baddeley, & Mishkin, 2006). The concept of ‘levels of processing’ was proposed by Craik and Lockhart (1972), who claimed that remembering information depended not only on having attended to it during its occurrence or having rehearsed it after its occurrence, but also on how deeply it was processed. Their proposal included the following notions: (a) Processing is the analysis of information that underlies perception and comprehension; (b) There is a hierarchy of levels of analysis, running from early analyses of sensory and surface features to later analyses of semantic and conceptual features; (c) Analysis of meaning requires more attention than analysis of sensory features, and, therefore, the later levels of analysis are “deeper” levels of processing; and (d) Semantic deeper processing is associated with stronger and longer-lasting memory traces. Craik and Tulving (1975) found empirical evidence for the effects of levels of processing on both incidental and intentional memory performance.

Empirical studies have also investigated the effect of levels of processing on experiences of remembering and knowing, and not only on memory performance, by
using the remember-know procedure\(^3\) with different types of items, such as L1 words in the visual mode (Gardiner, 1988; Rajaram, 1993; Java, 1994; Gardiner, Gawlik & Richardson-Klavehn, 1994; Richardson-Klavehn & Gardiner, 1995; Gardiner, Java & Richardson-Klavehn, 1996); L1 words in the auditory mode (Rajaram, 1993; Toth, 1996; Karayianni and Gardiner, 2003); pictures (Rajaram, 1996; Gardiner, Gregg, Mashru & Thaman, 2001; Konstantinou & Gardiner, 2005); and melodies and songs (Gardiner, Java & Richardson-Klavehn, 1996; Gardiner & Radomski, 1999). Overall, these studies have found support for two main hypotheses: (1) Fairly minimal encoding conditions (shallow processing) are sufficient for items to be registered in semantic memory, and/or to be recognized with an experience of knowing; and (2) Optimal encoding conditions (deep processing) are necessary for items to be registered in episodic memory, and/or to be recognized with an experience of remembering.

Tables 1-4 below present remember-know studies that have examined the effects of levels of processing (LOP) on recognition memory of L1 words in the visual or auditory modes. These studies present the following operationalizations of deep and shallow processing: (a) encoding while carrying out semantic versus a graphemic/phonetic tasks (e.g., studies on Table 1); (b) encoding while completing generating versus reading tasks (e.g., studies on Table 2); (c) encoding while carrying out undivided attention versus divided attention tasks (e.g., studies on Table 3); and (d)

\(^3\) Most studies that address the effects of levels of processing have employed the remember-know procedure. However, there are also studies that have used other methods, such as PDP and ROC. For references of these studies see Yonelinas (2002).
encoding of words versus non-words (e.g., studies on Table 4). Thus, deep processing is related to conceptual or semantic processing, amount of attention, and elaborative processing. Most studies use a within-subject design so that all participants complete both types of task. Number of items for each condition ranges between 15 and 100 (e.g., Yonelinas, 2001), and variables such as number of letters or syllables, frequency, and, to a lesser extent, word class and word imageability, are controlled. The retention interval ranges from a few minutes in most cases to one week. Findings of these studies indicate that deep processing gives rise to significantly more remember responses than know responses. In contrast, shallow processing does not have a significant effect on either remember or know responses.

---

4 A narrow definition of LOP includes only semantic versus perceptual processing. Here I use the broad definition used in current research, which extends the notion of LOP to amount of attention and elaboration during processing.

5 The remember-know task is usually preceded by unrelated filler activities that take between 3-25 minutes, so as to not allow participants to rehearse target items, or keep them active in working memory.
Table 1. Remember-know studies addressing effects of LOP on recognition memory of L1 words

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Deep processing as encoding in a semantic task</th>
<th>Shallow processing as encoding in a perceptual task</th>
<th>Memory experience</th>
<th>Retention interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gardiner (1988, Exp.1)</td>
<td>18 college students, UK</td>
<td>Writing a semantic associate of a word</td>
<td>Writing a rhyming word</td>
<td>R / K responses</td>
<td>10 min.</td>
</tr>
<tr>
<td>Rajaram (1993, Exp.1)</td>
<td>16 undergrad students in UK</td>
<td>Writing a semantic associate of a word (visual + auditory modes)</td>
<td>Writing a rhyming word (visual + auditory modes)</td>
<td>R / K responses</td>
<td>1 hour</td>
</tr>
<tr>
<td>Gregg &amp; Gardiner (1994, Exp.1)</td>
<td>32 college students in UK</td>
<td>Indicating by a tick if a word represents a four-footed animal (visual + auditory)</td>
<td>Indicating by a tick if a word could be read or heard distinctly or not</td>
<td>R / K responses</td>
<td>10 min.</td>
</tr>
<tr>
<td>Perfect et al. (1995, Exp. 2A)</td>
<td>48 subjects: 24 old &amp; 24 young</td>
<td>Forming an image of a word + rating ease of imagery generation (visual mode)</td>
<td>Counting the number of vowels for each word and writing it down (visual mode)</td>
<td>R / K responses</td>
<td>25 min.</td>
</tr>
<tr>
<td>Gardiner et al., (1996a, Exp.1, 2)</td>
<td>16 undergrads (Exp. 1), 20 young adults (Exp. 2)</td>
<td>Writing a semantic associate of the word (visual mode)</td>
<td>Writing two letters that do not occur in the word (visual mode)</td>
<td>R / K / G responses</td>
<td>1 day (Exp. 1) 1 week (Exp. 2)</td>
</tr>
<tr>
<td>Java et al. (1997, Exp.2)</td>
<td>30 undergrad students in UK</td>
<td>Writing a semantic associate of a word (visual)</td>
<td>Writing a rhyming word (visual mode)</td>
<td>R / K responses</td>
<td>15 min.</td>
</tr>
<tr>
<td>Gardiner et al. (1999, Exp.1)</td>
<td>32 undergrad students in UK</td>
<td>Rating the ease of generating semantic associates on a 5-pt scale</td>
<td>Rating the ease of generating rhyming words on a 5-pt scale</td>
<td>R / K / G responses</td>
<td>A few minutes</td>
</tr>
<tr>
<td>Kho et al. (2000, Exp.1)</td>
<td>32 undergrad students in US</td>
<td>Generating an associate of the word (visual mode)</td>
<td>Generating 2 vowels not present in the stuffy word</td>
<td>R / K responses</td>
<td>1 week</td>
</tr>
<tr>
<td>Yonelinas (2001, Exp.3)</td>
<td>54 students in a psychology course in USA</td>
<td>Rating the pleasantness of the word on a 4-point scale (auditory mode)</td>
<td>Counting the number of syllables in each word (auditory mode)</td>
<td>R / K resp. (+PDP, ROC)</td>
<td>Immediate</td>
</tr>
<tr>
<td>Richards-Klavehn et al., 2002, Exp. 2A)</td>
<td>48 university students in UK</td>
<td>Rating the pleasantness of a word on a scale from 1 to 7 (visual mode)</td>
<td>Judging whether a word shared any vowels with the previous word (visual)</td>
<td>R / K responses (+ PDP)</td>
<td>Immediate</td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Deep processing as encoding in a generating task</td>
<td>Shallow processing as encoding in reading aloud</td>
<td>Memory experience</td>
<td>Retention interval</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>---------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Gardiner et al. (2006, Exp. 1)</td>
<td>An adult with amnesia + control group</td>
<td>Rating words for their pleasantness (visual mode)</td>
<td>Counting the number of syllables in a word</td>
<td>R / K responses</td>
<td>Immediate</td>
</tr>
<tr>
<td>Gardiner et al. (2006, Exp. 2)</td>
<td>An adult with amnesia &amp; control group</td>
<td>Rating words for their imagery (visual mode)</td>
<td>Rating words for their length (visual mode)</td>
<td>R / K responses</td>
<td>Immediate</td>
</tr>
<tr>
<td>Table 2. Studies addressing the generation effect on memory experiences</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gardiner (1988, Exp. 2)</td>
<td>40 undergrad students in UK</td>
<td>Generating a word in the context of a given rule (e.g., “opposites”). The first letter is provided (e.g., hot-_.?)</td>
<td>Reading aloud a word (e.g., hot-cold)</td>
<td>R / K responses</td>
<td>1 h</td>
</tr>
<tr>
<td>Wippich (1992, Exp. 1)</td>
<td>48 undergrad students in Germany</td>
<td>Exchanging 2 letters underlined in a word + generating resulting word</td>
<td>Reading aloud a word</td>
<td>R / K responses</td>
<td>A few minutes</td>
</tr>
<tr>
<td>Gardiner et al. (1996, Exp. 3)</td>
<td>18 university students in UK</td>
<td>Generating a word from a semantic context</td>
<td>Reading aloud a word without any context</td>
<td>R / K / G responses</td>
<td>1 week</td>
</tr>
<tr>
<td>Gardiner et al. (1999, Exp. 2)</td>
<td>40 undergraduates in UK</td>
<td>Generating a word, given a previous definition and the initial letter (e.g., Object or area to aim at: T___)</td>
<td>Generating a word (e.g., Target)</td>
<td>R / K / G responses</td>
<td>3-7 days (depending on availability)</td>
</tr>
<tr>
<td>Curran &amp; Hildebrandt (1999)</td>
<td>16 healthy volunteers, moderate social drinkers</td>
<td>Generating and saying aloud a word in the context of a given rule (e.g., “opposites”). The first letter is provided.</td>
<td>Reading aloud a word</td>
<td>R / K responses</td>
<td>A few minutes</td>
</tr>
</tbody>
</table>
Table 3. Studies addressing the effect of undivided vs. divided attention during encoding on memory experiences

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Deep processing as encoding in an undivided attention task</th>
<th>Shallow processing as encoding in a divided attention task</th>
<th>Memory experience</th>
<th>Retention interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gardiner &amp; Parkin (1990)</td>
<td>54 members of a laboratory (grad students &amp; staff)</td>
<td>Attending to the words visually presented and trying to memorize them</td>
<td>Attending and studying words while detecting tones played on an audiotape at different speed (slow/fast)</td>
<td>R / K responses</td>
<td>10 minutes</td>
</tr>
<tr>
<td>Yonelinas (2001, Exp.1)</td>
<td>19 undergrads enrolled in a psychology course in USA</td>
<td>Reading aloud each word presented on a screen and try to memorize it</td>
<td>Reading aloud each word along with 2 random numbers between 1 and 9 to the left. When a third random number appeared on the screen, participants had to press a yes key if the final number was between the first 2 numbers in value.</td>
<td>R / K responses (+ confidenc e ratings for ROC)</td>
<td>Immediate</td>
</tr>
<tr>
<td>Yonelinas (2001, Exp. 2a)</td>
<td>18 undergrads enrolled in a psychology course in USA</td>
<td>Same as Yonelinas (2001, Exp. 1)</td>
<td>Same as Yonelinas (2001, Exp. 1)</td>
<td>R responses + familiarity rating on a 6-point scale</td>
<td>Immediate</td>
</tr>
<tr>
<td>Yonelinas (2001, Exp 2b, 2c)</td>
<td>18 undergrads enrolled in a psychology course in USA</td>
<td>Same as Yonelinas (2001, Exp. 1)</td>
<td>Exp. (2b): Making a “less than or equal to” judgment about 2 digits while studying the words (rather than the “is it between these 2 values” judgment in previous experiments); Exp. (2c): Pressing the space bar whenever 3 odd numbers in a row appeared next to the word presented</td>
<td>R responses, + familiarity rating on a 6-point scale</td>
<td>Immediate</td>
</tr>
<tr>
<td>Mangels et al. (2001)</td>
<td>20 young subjects (28 yrs), native English speakers, w/o neurological disorder</td>
<td>Memorizing written words</td>
<td>Memorizing written words and identifying low/medium/high pitch tones (2 different degrees of difficulty)</td>
<td>R / K responses (+ ERPs)</td>
<td>A few minutes</td>
</tr>
</tbody>
</table>
Table 4. Studies addressing the effect of encoding words vs. non-words on memory experiences

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Deep processing as encoding of words</th>
<th>Shallow processing as encoding of non-words</th>
<th>Memory experience</th>
<th>Retention interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gardiner &amp; Java (1990, Exp. 2)</td>
<td>20 undergrads in UK</td>
<td>Memorizing written words</td>
<td>Memorizing written non-words</td>
<td>R / K responses</td>
<td>1 day</td>
</tr>
<tr>
<td>Rajaram et al., (2002, Exp. 1)</td>
<td>20 undergraduates in USA</td>
<td>Memorizing written words</td>
<td>Memorizing written non-words</td>
<td>R / K responses</td>
<td>15 min.</td>
</tr>
<tr>
<td>Karayianni &amp; Gardiner (2003)</td>
<td>16-32 undergrads without knowledge of Greek</td>
<td>Listening twice to spoken words recorded on a CD</td>
<td>Listening twice to spoken non-words recorded on a CD</td>
<td>R / K responses</td>
<td>10 min.</td>
</tr>
</tbody>
</table>

In the last years, however, researchers have acknowledged that the view of semantic and perceptual processing as deep and shallow processing respectively is not pertinent, since both conceptual and perceptual features can be processed with different degrees of depth (cf. Baddeley, 2002, and Craik, 2002, for a review of the levels of processing framework). Currently, the notion of depth of processing is related to ‘degree of elaboration’, ‘degree of consciousness’ or ‘level of awareness’ rather than to semantic processing. Gardiner and Richardson-Klavehn (2000) establish a relationship between these notions and memory: “All that is necessary for encoding into the semantic system is some initial awareness of the events, however fleeting. In contrast, encoding into episodic memory must depend on greater conscious elaboration of the events” (p. 234). According to this view, findings would give support for the coordination hypothesis, which posits a relationship between attention and awareness at encoding, on the one hand, and memory awareness at retrieval, on the other hand. However, the
operationalizations used for the notions of ‘degree of elaboration’, ‘degree of consciousness’ and ‘level of awareness’ in the empirical studies reviewed (e.g., encoding while completing generating versus reading tasks, undivided attention versus divided attention tasks, and encoding of words versus non-words6) present minimally the following methodological issues: (a) ‘deep’ and ‘shallow’ are not defined independently, that is, the definition of one level depends on the definition of the other level (e.g., memorizing real words is assumed to induce deep processing but only when this task is compared to memorizing non-words); and (b) process measures such as think-aloud protocols are not employed to ensure that participants’ performances are representative of their different experimental conditions (e.g., it is assumed that they pay attention in a task with no divided attention requirements). Therefore, most studies investigate the effects of different types of tasks or items that are assumed to reflect a type of processing but they do not attempt to measure the type of processing itself. For this reason, Craik (2002) has recently pointed out that current operationalizations of depth of processing are not completely satisfactory. According to the researcher, the main limitation of the levels of processing framework is the absence of an objective index of depth of processing.

With respect to measures of memory experiences, only three studies (Gardiner et al., 1996; Gardiner et al., 1999; Gardiner et al., 2006) include guess responses to control for the potential confounding of knowing with guessing. In order to check responses’

6 Craik (2002) has defined “elaboration” as “the degree to which each type of processing has been enriched during encoding” (p. 307). Mandler (2002) indicates that elaboration entails assimilating a target unit into existing mental meaningful structures: “Any deep analysis at encoding is a process of finding structures (meanings) into which the target item fits. There is no meaning to be extracted, meaning is in the structure in which the item is embedded” (p. 337). In this sense, non-words cannot be related to any meaningful structure, and, therefore, they provide less opportunities than words do for meaningful, elaborative encoding (Richardson-Klavehn, Gardiner, & Ramponi, 2002).
validity, participants are usually asked to explain a random selection of responses after taking the test; this procedure, however, can be questioned in light of two issues: (a) the ability to accurately distinguish between experiences may vary from item to item, and (b) the memory of the experience that led to choose a response may be weak after taking the test. Moreover, all studies present participants with lists of words that are not contextualized. Because the experience of remembering entails the recollection of contextual details, it seems that the design used might not be the most appropriate to facilitate experiences of remembering. Although the remember-know paradigm has certainly proved to be very useful to measure memory experiences in cognitive psychology, there are certain issues that could be overcome to improve the method and obtain more robust results, especially with investigating natural occurring languages.

Finally, most studies reviewed in this survey have used a short retention interval. Studies outside the levels of processing framework suggest that know responses may lead to lack of recognition after one week, while remember responses become know responses over time (Knowlton & Squire, 1995; Conway et al, 1997). This phenomenon is referred to as the ‘remember-to-know-shift’. Because the effect of retention interval may have important implications for learning, studies investigating depth of processing should also address this time variable.

As has been mentioned earlier, several models in cognitive psychology have proposed that remembering has a stronger impact on learning than knowing. In this line, Craik (2002) relates depth of processing and memory systems to ‘levels of knowledge representation’. Higher levels of knowledge are usually characterized by generality and
abstraction, while lower levels are characterized by knowledge of specific details. This has made the researcher question why episodic memory (i.e., memory of specific contextual details), that would be a priori associated with lower levels of knowledge, should be related to deep processing and longer-lasting recognition, and why it is more difficult to be accessed than semantic memory. Craik (2002) points out that deep processing and episodic memory cannot be related to lower levels of knowledge. According to him, “deep processing provides the schematic context within which episodic details are related to each other and to more abstract representations” (p. 313). Thus, episodic memories may play an important role in how abstract representations (i.e., higher levels of knowledge) are initially formed.

The idea that episodic memory is linked to higher levels of knowledge is also consistent with Tulving’s (1985) proposal that information is encoded into episodic memory only if it is also encoded into semantic memory. In this sense, episodic memory results in a “richer” encoding that can lead to higher levels of performance over time. Furthermore, the remember-to-know shift found in several studies would also support the idea that episodic memory is related to higher levels of knowledge, since memories of specific contextual details seem to become semantic memories over time. Findings from Conway et al.’s (1997) study give evidence for this relationship between episodic memory and learning. Conway et al. (1997) conducted a large-scale naturalistic study, where first-year psychology students were given multiple-choice tests following each of three research methods courses and four lecture courses they were enrolled in. Participants in one lecture course (N=64) were retested. They were asked to recognize
course content by selecting an answer to multiple choice questions; in addition, they had to indicate if an answer had been selected because: (a) they remembered a learning episode in which they had encountered the content (remember response); (b) they just knew that this was the correct answer, had a strong feeling of knowing and did not remember a learning episode (just know response); (c) they neither remembered the learning episode or knew the answer, but felt the chosen alternative was more familiar (familiarity response); (d) they felt they were guessing, for example, choosing the alternative that looked least unlikely (guess response). Results from an ANOVA on correct answers with response and time as within-subject factors yielded a significant interaction between both variables. Separate analyses revealed that in the first posttest participants produced significantly more remember responses, while in the second posttest remember responses significantly decreased and know responses significantly increased. Differences for familiar and guess responses were not significant. This study showed a remember-to-know shift, and led the researchers to conclude that episodic memory may play an important role in learning. However, further research is needed to consolidate this finding and generalize it to different areas of learning.

**Empirical Studies in SLA**

To my knowledge, only three unpublished studies (Peckham, 2000; Martínez-Fernández, 2009a, 2009b) have applied the remember-know paradigm to investigate cognitive experiences in SLA. In his dissertation, Peckham (2000) conducted a classroom-based study to investigate the impact of L2 instruction on noticing, as well as
the relationship between noticing and learning of both L2 vocabulary and grammar. In this study, the remember-know procedure was employed to measure noticing, which was operationalized as a remember response. Data from 26 and 14 participants were collected in two rounds respectively, with 13 participants being included in both rounds. Intact classes were assigned to two groups, varying in whether they received instruction on grammar and exposure to vocabulary or instruction on vocabulary and exposure to grammar. Participants took a pretest, attended two sessions of instruction, in which participants read a text containing all target items (12 abstract words and 6 grammatical structures), and completed three posttests, each including noticing, grammar, and vocabulary tests.

The noticing test consisted of two parts: reading of a short text containing both instructed and non-instructed items, and the remember-know test. In the remember-know test, participants were given pairs of items, and were asked to recognize which item was in the text they had just read, and to give a remember, know or guess response. In order to get accurate responses on this test, the researcher trained the participants during the instruction sessions, informed them that they would have to take this type of test later on, distributed written instructions, and conducted individual interviews after the test. Finally, the grammar test consisted of a sentence completion task and a grammatical judgment task; in the vocabulary test, participants were asked to provide the translation or definition of the target items.

Because testing times were different between and within rounds of data collection, and target items and type of pretest were changed from one round to the other,
no comparisons could be made either across the rounds of data, or across the testing times within each round. Data were submitted to several paired t-tests and correlations. Although the majority of results were not statistically significant, the analysis of data from the first round on the first posttest showed the following findings: (a) grammar instruction had a significant effect on noticing of grammatical structures when compared to grammar exposure, which led the researcher to conclude that instruction can make aspects of the input more salient, and lead to noticing; however, the difference between noticing of instructed grammatical and non-instructed grammatical items within the grammar instruction group was not significant, and while the effect of instruction on subsequent posttests decreased, the effect of exposure increased; (b) noticing of vocabulary was significantly higher in the vocabulary exposure condition than noticing of grammar in the grammar exposure condition, suggesting that vocabulary may be easier to notice, and that grammar may benefit from instruction to a greater extent; and (c) no significant correlation between noticing and learning was found, although the researcher attributed this lack of significance to the unreliability of the data, and claimed that noticing and learning at least co-occurred in most of the cases.

Beyond the issues related to the sample size and the methodological differences between rounds of data collection, this study presents other methodological limitations, such as including the same participants in both rounds, including participants with prior knowledge of some items, lack of pretest for non-instructed items that were targeted in
the noticing test, lack of control of outside exposure,⁷ and exposure to and information about the noticing test during the instruction sessions, among others. More importantly, this study fails to distinguish concepts such as deep processing, noticing and memory, and to provide adequate operationalizations. On the one hand, the researcher seems to match noticing with deep processing, and assumes that both are reflected in remember responses, while shallow processing is assumed to be reflected in know responses; yet, in the discussion the researcher considers that noticing might be reflected both in remember and know responses. On the other hand, the study does not provide a distinction between awareness and memory, since noticing is operationalized as remembering. Noticing, however, takes place when participants are processing some material. Therefore, learners may notice a particular item, and may not be able to remember it. Noticing might certainly have an effect on remembering, but the relationship between awareness and experiences of recognition memory in SLA still needs to be investigated.

Martínez-Fernández (2009b) investigated the effect of task complexity in lexical tasks during reading, as well as the effect of thinking aloud while reading, on experiences of remembering and knowing, L2 vocabulary development, and text comprehension. Task complexity was defined by the reasoning demands required by the tasks. In the experimental conditions, participants read a text containing eight blanks, and had to choose one of two unfamiliar words, which were bolded and glossed at the margin, in order to fill in each blank. In the ‘high reasoning’ conditions, both options given for each

⁷ The researcher indicates that some participants of one group were overheard discussing what the opposite group was learning, which clearly might have led them to focus on the area where they did not receive instruction but were tested on.
blank were plausible but only one was correct based on specific contextual clues. In the ‘low reasoning’ conditions, one of the two options given was clearly incorrect because it had the opposite lexical feature ([+/- concrete]) to that selected by the previous word. Participants were one hundred forty-seven American college students enrolled in a fourth-semester Spanish course. They completed a pretest, and were randomly assigned to one of five groups, four groups varying in the degree of reasoning required by the lexical task ([+R] or [-R]) and the presence or absence of thinking aloud while reading ([+TA] or [-TA]), and one control group, in which participants were asked to read the text without glosses or blanks and think aloud. Immediately after the treatment, participants completed a text comprehension test, followed by a remember-know task, and two vocabulary tests. A one-week delayed posttest measured memory experiences and vocabulary retention.

Before completing the remember-know task, participants were given written and oral instructions consisting of detailed definitions of remembering, knowing and guessing, and examples in real-life situations. In addition, four participants who were excluded from the final sample were asked to think aloud while completing this test. Items were presented one by one with a PowerPoint presentation, and participants provided their answers in an answer sheet. As in cognitive psychology studies, they had to decide whether an item had been previously presented in the text (old item) or not (new item), and select a remember, know or guess response for old items.

The statistical analysis of data on memory experiences from a final pool of 71 participants showed that (a) the [+R] groups remembered significantly more words than
the control group, while there was no significant difference between the [-R] group and 
the control group; (b) thinking aloud did not have an effect on the immediate posttests, 
but the [+R, +TA] group remembered significantly more targeted words in the delayed 
posttest than the [+R, -TA], which suggests that thinking aloud while performing a 
complex task might help maintain episodic memories over a period of a week; and (c) the 
 [+R, -TA] experienced a significant decrease of remember responses over time, together 
with a significant increase of know responses; this result is consistent with findings in 
cognitive psychology studies that show a significant loss of episodic memories over time, 
and a remember-to-know shift. Finally, the [+R] groups also produced and recognized the 
meaning of significantly more words than the control group, while there was no 
significant difference between the [-R] group and the control group. This finding 
indicates that there might be a relationship between memory experiences and L2 
vocabulary development, although the study did not address this correlation.

The analysis of the think-aloud protocols of four participants who thought aloud 
while completing the remember-know task provided rich qualitative data to support the 
different experiences of remembering and knowing. This analysis revealed that in some 
cases guess responses were assigned to experiences of knowing, and know responses to 
experiences of remembering. Concurrent verbalizations proved be an efficient method to 
ensure the validity of participants’ responses. However, it is important to note that in this 
study the researcher did not check responses’ validity since participants in the final 
sample did not think aloud while completing the remember-know task, and they were not 
asked to provide any explanation for their responses once they had completed the test.
Finally, Martínez-Fernández (2009a) presented findings of the previous study regarding the effect of thinking aloud while reading and completing a complex task. In addition, the researcher compared memory experiences triggered by both targeted items and non-targeted items that were present in the experimental text but were not deleted, bolded or glossed, since they were frequent words that participants were assumed to be familiar with. Participants (N= 63) were randomly assigned to one of two groups varying in whether they were asked to think aloud or not. The statistical analysis from a final sample of 35 participants yielded the following results: (a) both groups produced significantly more remember responses for non-targeted items than for targeted items on both posttests, and (b) the [+TA] group produced significantly more remember responses than the [-TA] group for non-targeted items on both posttests. These findings indicate that prior knowledge may have an effect on episodic memory, and that the effect of thinking aloud may interact with prior knowledge to help create and maintain episodic memories of familiar words. These results may be related to those found in previous research in cognitive psychology, which show that L1 words (i.e., familiar words) lead to significantly more remember responses than non-words (i.e., words that participants have never been exposed to before the experiment). Therefore, the difficulty to commit unfamiliar words to episodic memory might explain to some extent the challenges of second language learning. This study, however, did not address the correlation between episodic memory and L2 development. A relationship between memory experiences and

---

8 Although these words had not been included in the vocabulary pretest, participants had been repeatedly exposed to them in Spanish classes.
learning may also have important implications for research on type of task, as task types that lead to enhanced episodic memory might be considered more beneficial than others.

Conclusion

Previous research in cognitive psychology indicates that deep processing and awareness of L1 words has an effect on remembering. Although SLA research has empirically identified different levels of awareness (e.g., Leow, 1997a; Martínez-Fernández, 2008a; Rosa & Leow, 2004; Rosa & O’Neill, 1999; Sachs & Suh, 2007), and also suggests a relationship between awareness and episodic long-term memory (e.g., Robinson, 1995b; N. Ellis, 2005), this relationship has never been empirically tested. To date, only three unpublished studies have measured experiences of remembering and knowing in SLA. Peckham (2000) applied the remember-know procedure to L2 development but this study failed to distinguish awareness from memory experiences, and Martínez-Fernández (2009a, 2009b) did not address the relationship between these cognitive experiences.

Future studies, therefore, should investigate the relationship between levels of awareness and memory experiences in SLA, and overcome a number of issues found in the literature regarding the operationalization of these constructs. First, the measures used for depth of processing in cognitive psychology (i.e., undivided/divided attention tasks, generating/reading tasks, words/non-words encoding) are not satisfactory (Craik, 2002). Research studies usually confound type of processing and type of task. While task demands may have an effect on memory experiences and other cognitive processes (cf.
Robinson, 1995b, 2003), this effect should not be attributed to a certain type of processing if processes are not measured. Concurrent verbal protocols, which have the advantage of measuring processes directly at the same time they are taking place, may be a more appropriate measure, as long as potential reactivity is controlled. Moreover, the combination of concurrent verbalizations with the remember-know procedure may allow researchers to answer a number of questions that remain controversial in the cognitive psychology field. Specifically, concurrent verbalizations may be useful to determine what experiences accompany memory retrieval, what characteristics are associated with these experiences, and to control for response criterion. In addition, an online measure seems more appropriate than an offline measure to check responses’ validity, since memory experiences may easily decay over a short period of time.

Furthermore, because both awareness and type of task or type of exposure might have an effect on memory experiences and L2 development, future studies should investigate and distinguish the effects of both variables. Such research would help expand the remember-know framework to include the relationship between recognition memory experiences and learning, which has not received enough attention either in cognitive psychology or in SLA. Martínez-Fernández (2009a, 2009b) investigated the effect of task complexity, and found that more complex tasks might lead to higher episodic memory and vocabulary development than less complex tasks. However, the researcher did not address the correlation between memory experiences and learning.

Finally, Peckham (2000) found differences between lexical and grammatical targeted items in the remember-know task, concluding that the effect of exposure to
vocabulary on episodic memory of lexical items was higher than the effect of exposure to grammar on episodic memory of grammatical items. However, the fact that six different grammatical structures were tested in this study might have had an effect on these results. This study did not find any correlation between responses in the remember-know task and vocabulary and grammar learning, but the methodological limitations mentioned above do not allow us to come to any conclusion. Overall, researchers seem to agree on the important role of declarative memory on both L2 vocabulary and L2 grammar, even though they present different predictions regarding the weight of declarative memory in vocabulary and grammar learning (e.g., Ullman, 2001; N. Ellis, 2001). In spite of the importance of declarative memory in L2 development, it is not clear yet whether (a) episodic and semantic memory play different roles in L2 development, and (b) whether episodic and semantic memory play the same role in grammar as in vocabulary. Research studies in the cognitive psychology field have mainly investigated memory of lexical items in the first language, and therefore these findings cannot be generalized to other areas.

To summarize, future studies in SLA should expand previous findings by investigating experiences of recognition memory in relation to minimally the following research areas: (a) levels of awareness, (b) type of task/exposure, and (c) type of linguistic item (i.e., grammar versus vocabulary). The present study aims to undertake this research within an attentional framework.
Levels of Awareness in SLA

An increasing number of empirical studies in SLA have found evidence for different levels of awareness as measured by concurrent verbalizations (e.g., think-aloud protocols), and their effect on L2 development (e.g., Leow, 1997a; Martínez-Fernández, 2008a; Rosa & O’Neill, 1999; Rosa & Leow, 2004; Sachs & Suh, 2007). The measure of awareness, however, is still controversial both in the cognitive psychology field and in SLA. Some researchers have argued that concurrent verbalizations produced while completing a task may alter the cognitive processes involved in task completion, as well as the outcome of such processes (e.g., Ericsson & Simon, 1993). The sections below present (a) a brief review of SLA literature on the operationalization of awareness, and empirical studies addressing methodological issues, (b) a review of empirical studies on reported levels of awareness of grammatical items (Camblor, 2006; Leow, 1997a; Rosa & O’Neill, 1999; Rosa & Leow, 2004), and (c) a review of empirical studies on reported levels of awareness of lexical items (Bowles, 2004; Martínez-Fernández, 2008a; Guidi, 2009).

Operationalization of Awareness

Schmidt (1990 and elsewhere) suggests that awareness can be operationally defined as the ability for verbal report during or after exposure. In the same line, Robinson (2003) makes a distinction between the operationalization of deep processing and that of noticing (i.e., awareness at a low level). The researcher establishes a relationship between these two constructs, as he claims that noticing emerges in short-
term memory out of elaborative and conceptually driven processing (i.e., deep processing). In his view, however, deep processing can be operationalized by task demands, as has been usually done in the cognitive psychology field, while the measure of noticing would require learners’ reports. Although different types of tasks may promote different levels of processing, this operationalization of depth of processing should be taken with caution, as has been mentioned earlier, since processing depends primarily on what the learner does when interacting with the material to be learned (regardless the task that he or she completes). In any case, it is important to note that (a) the notion of levels of awareness in SLA and that of levels of processing developed in cognitive psychology are operationalized in different ways, and (b) the notions of levels of awareness and levels of processing are related but are not equivalent: deep processing may be related to both noticing (awareness at a low level) and understanding (awareness at a high level), while shallow processing may not be not related to awareness.

Both Schmidt (1990 and elsewhere) and Robinson (2003) agree on the idea that the operationalization of awareness requires learners’ reports. Two main issues, however, make the attempt to operationalize and measure awareness problematic: (a) the rapidity of a learner's subjective experience of cognitive registration, and (b) the potential inability to verbalize one's awareness (Schmidt, 1995; Leow, 1997a; Leow, 2000).

Different procedures that have been used to measure awareness in SLA include diaries, offline questionnaires, stimulated recall, immediate retrospective verbal reports and online think-aloud protocols, among others (cf. Egi, 2004 for a review of methods). Think-aloud protocols have the advantage of providing reports of experiences and
processes at the same time they are taking place, thus avoiding the memory decay problems associated with other measures. However, there is some concern about the validity of this measure. Thinking aloud while completing a task may alter the primary process, and, therefore, the outcome of that process, on the one hand, and may cause a delay in task completion time (i.e., a latency effect), on the other hand (Ericsson & Simon, 1993). This issue is referred to in the literature as ‘reactivity’.

A growing number of studies have recently addressed the issue of reactivity in different areas of SLA. In their pioneering study, Leow and Morgan-Short (2004) examined the effect of non-metalinguistic L1 verbal reports of seventy-seven learners of Spanish during incidental reading on text comprehension, intake, and controlled written production of a morphological structure. Findings of this study show a latency effect but no reactivity for accuracy on any of the tests. Bowles and Leow (2005) expanded the scope of this investigation to examine the potential reactivity of both non-metalinguistic and metalinguistic L1 verbal reports (i.e., reports explaining the thought processes) of forty-five learners of Spanish on text comprehension and item and system learning of a syntactic structure. Results of this study show that neither type of verbalization significantly affected text comprehension or written production of both old and new exemplars of the targeted structure. However, metalinguistic verbalization did appear to cause a significant decrease in text comprehension when compared to non-metalinguistic verbalization.

The lack of reactivity of non-metalinguistic L1 think-aloud protocols on grammatical development has been found in relation to a variety of types of tasks in
subsequent studies. Bowles (2008) compared non-metalinguistic and metalinguistic L1 verbalizations of 194 learners of Spanish produced while completing a problem-solving task, and investigated the potential interaction of thinking aloud with more and less explicit feedback. This study showed that thinking aloud in the metalinguistic conditions caused a delay on task completion time, as well as reactivity on production of a syntactic structure. Simply thinking aloud (i.e., non-metalinguistic verbal reports), however, did not hinder grammatical development in any of the conditions receiving more or less explicit feedback. No reactivity for non-metalinguistic L1 think-aloud protocols was found also by Sachs and Suh (2007), who used a picture description task in computer-mediated communication to investigate the effect of textual enhancement of recasts on written production of a morphological structure by thirty learners of Korean. Finally, Guidi (2009) also reported lack of reactivity on both grammatical and lexical recognition and written production tests in a study that examined the effects of glossing and type of linguistic item on the reading comprehension, noticing and L2 learning experienced by sixty-five learners of Spanish.

Based on the idea that the effect of thinking aloud might depend on type of task, Medina (2008) investigated whether reactivity of concurrent verbalizations produced by eighty learners of Spanish was influenced by task complexity. In addition, this study examined whether working memory might influence reactivity. The results indicated an effect of thinking aloud on time on task in both more complex and less complex task conditions, but no reactivity for any condition on recognition and written production of a Spanish morphological structure. Moreover, this lack of reactivity on accuracy was not
related to participants’ working memory (cf. Goo, 2010 below for different results on the relationship between reactivity and working memory). Similar findings on reactivity and task complexity were presented by Martínez-Fernández (2008b) from a study that compared two lexical tasks with different degree of cognitive complexity in both think-aloud and silent conditions. The results of this study showed a delay on task completion time in the think-aloud conditions but a lack of reactivity on the text comprehension and vocabulary development of seventy-one learners of Spanish. Finally, Yoshida (2008) investigated the interaction between reactivity and cognitive complexity of reading tasks completed by sixty-four Japanese learners of English. Findings of this study showed that thinking aloud slowed down performance on task, but did not affect text comprehension in either more complex or less complex conditions.

Currently, only three studies indicate that thinking aloud may affect L2 performance (Goo, 2010; Sanz, Lin, Lado, Bowden, & Stafford, 2009; Sachs & Polio, 2007). Goo (2010) investigated working memory capacity and performance of 42 English-speaking learners of Spanish on reading comprehension and written production under think-aloud versus non-think-aloud conditions. Results led the researcher to conclude that thinking aloud might negatively affect learner performance on reading comprehension. Interestingly, the reactive effects seemed to have occurred among the high working memory capacity learners and not among those with low working memory capacity. Thus, this study shows that reactivity may depend on individual differences in working memory capacity.
Sanz et al. (2009) found evidence for positive reactivity, and concluded that non-metalinguistic L1 think-aloud protocols might have a facilitating effect on grammatical development with certain pedagogical treatments. In this study, the researchers report findings of two experiments where 24 naïve learners of Latin received a computerized treatment that delivered a grammar lesson, practice, and feedback on Latin grammar in the first experiment, and only practice and feedback in the second one. The results showed that thinking aloud did not affect reaction time on subsequent tests, but the less explicit condition induced positive reactivity on grammatical development, in contrast to the more explicit condition, which did not induce reactivity, either positive or negative. The effect size of the significant effect found, however, was not reported. Finally, Sachs and Polio (2007) used non-metalinguistic L2 think-aloud protocols to investigate the effects of reformulations versus written error correction on awareness and L2 grammatical accuracy of fifteen ESL students. The results of this study indicated that thinking aloud had a negative effect in the reformulation condition. This finding is not surprising, since thinking aloud in a second language requires a greater cognitive effort than thinking aloud in the first language. In addition, it should be noted that the sample size in these last two studies that found reactivity (i.e., Sanz et al., 2009; Sachs & Polio, 2007) is much smaller than that of most studies indicating a lack of reactivity on L2 development.

Finally, Martínez-Fernández (2009a, 2009b) examined the effect of thinking aloud on recognition memory experiences. As has been described in the previous section, findings of this study suggest that thinking aloud while completing complex tasks may
help maintain the episodic memory of lexical items over a period of a week, while no reactivity is found in the less complex task conditions. Yet, this effect on memory experiences did not lead to reactivity on L2 vocabulary development.

Although the effects of thinking aloud need to be further investigated in relation to different types of treatments, assessment tasks, items, and individual differences, to date there is not enough evidence to conclude that concurrent verbalizations in the first language with no metalinguistic requirements affect L2 development. Nonetheless, every research design that incorporates this procedure should control for potential reactivity.

Beyond the issues related to reactivity, the use of online think-aloud protocols to measure awareness is still problematic, since learners may differ in their ability to verbalize their thoughts, and lack of verbalization does not imply lack of awareness. However, this limitation applies to all other procedures used to measure awareness, such as retrospective verbal reports or stimulated recalls. Moreover, in contrast to these procedures, concurrent verbalizations may capture learners’ subjective experiences with more precision, since these experiences are reported while they are taking place and not after a period of time. Therefore, in spite of the limitations that this measure certainly involves, it seems to be the most successful tool that we currently have to measure cognitive experiences.

**Empirical studies on reported levels of awareness of grammatical items**

As mentioned earlier, Schmidt (1990 and elsewhere) distinguished between awareness at a low level, ‘noticing’, and awareness at a high level, ‘understanding’,
which involves metalinguistic awareness. A number of empirical studies using think-aloud protocols have recently found evidence for the existence of levels of awareness and their differencial effects on grammatical development (Camblor, 2006; Leow, 1997a; Rosa & O’Neill, 1999; Rosa & Leow, 2004; Sachs & Suh, 2007). In Leow’s (1997a) study, twenty-eight adult beginning learners of Spanish were exposed to the 3rd. person preterit forms of Spanish in crossword puzzle, and were asked to think aloud while completing the task. Think-aloud protocols were coded to first establish learners’ noticing, before addressing the role of level of awareness in learners’ intake and accurate written production of the targeted forms. Noticing was operationalized as behavioral change accompanied by the verbalization of the targeted form. In addition, the protocols provided evidence of awareness at a higher level, understanding, that was operationalized as behavioral change accompanied by the verbalization of the morphological rule, and awareness at a lower level, that was operationalized as behavioral change without any type of verbalization. The results indicated that learners demonstrating awareness at a higher level significantly outperformed those showing awareness at lower level on a written recognition posttest, and, to a lesser extent, on a written production posttest.

Rosa and O’Neill (1999) also measured levels of awareness by using think-aloud protocols. In their study, sixty-seven native English speakers enrolled in a fourth-semester Spanish class were exposed to a syntactic form, Spanish contrary-to-fact conditional sentences in the past, and were assigned to five different conditions varying in the type of exposure: all groups completed ten multiple choice jigsaw puzzles, but differed in whether or not they were provided with grammatical explanation, directions
for rules search, directions to read a text, or directions to memorize. The think-aloud protocols revealed the following levels of awareness: ‘noticing’, operationalized as a verbal reference to the target structure without any mention of the rules, ‘understanding’, operationalized as a explicit formulation of the rule underlying the target structure, and ‘no verbal report’, operationalized as reading aloud the sentences without giving any verbal signal that either the verb in the target structure or the rule governing it had been cognitively registered. Chi-square analyses of frequencies of levels reported across the five experimental conditions indicated that (a) higher levels of awareness were related to more intake, and (b) although different levels of awareness were found in all types of exposure, formal instruction and directions for rules search were significantly related to understanding.

Rosa and Leow (2004) addressed the issue of levels of awareness measured by think-aloud protocols in a study in which one-hundred adult learners enrolled in a fifth-semester Spanish class were exposed to Spanish contrary to fact conditional sentences in the past. Participants were assigned to one of six conditions that differed in whether or not the exposure included a pretask with grammatical explanation and practice, task essentialness, implicit feedback, and explicit feedback. All groups were exposed to tenty-eight interactive LIBRA cards, each one containing a jigsaw puzzle. The results indicated (a) a relationship between higher levels of awareness and both the task essentialness conditions and the most explicit conditions, and (b) a relationship between higher levels of awareness and accurate recognition and written production of new items on both an immediate posttest and a delayed posttest administered three weeks later (although
participants who reached the highest level of awareness experimented a significant loss from the immediate posttest to the delayed posttest).

Finally, Camblor (2006) investigated whether different types of computerized feedback provided during writing were associated with different levels of awareness, and whether reported levels of awareness had a differential effect on controlled production of a morphological Spanish structure (gender and number noun-adjective agreement). Seventy-seven beginning learners of Spanish were randomly assigned to one of four conditions varying in the type of feedback received: explicit, implicit, interactive, and no feedback. The results indicated that type of feedback did not have an effect on levels of awareness, and that higher levels of awareness were significantly associated with higher performance on both old and new items, although this beneficial effect decreased over time.

In sum, SLA studies using think-aloud protocols reveal a significant relationship between reported levels of awareness and accurate recognition and production of both old and new grammatical items on immediate posttests.

**Empirical studies on reported levels of awareness of vocabulary items**

Reported levels of awareness have been investigated mainly in the area of grammatical development. However, Schmidt (1995) extended the notion of noticing to vocabulary development, and suggested that awareness at a lower level would involve noticing of few word aspects, while awareness at a high level would involve noticing of more word aspects (e.g., word form, meaning and syntactic features).
In the incidental L2 vocabulary learning strand, Bowles (2004) investigated whether glossing in both computer and paper-and-pen formats promoted awareness of the targeted words significantly more than an unglossed text. Fifty beginning learners of Spanish split into three groups: computer, non-computer and control conditions. Participants thought aloud while reading the text, answered 12 multiple-choice comprehension questions in English immediately after reading, and completed recognition and production vocabulary posttests. Noticing was operationalized as reading the glosses out loud or any comments about the targeted words (e.g., contextual guessing, commenting lack of knowledge of a word). Results showed that the gloss conditions outperformed the control group in comprehension and vocabulary posttests. In addition, the gloss conditions reported noticing of the targeted words significantly more than the control group, and no higher levels of awareness, related to higher-level processes such as hypothesis testing, were found for either the targeted words or other words in the text. The researcher concluded that (a) learners are less likely to use strategies like hypothesis testing with vocabulary words than with syntactic or morphological structures, and (b) glosses provide learners with all the necessary information for their goal (understanding the text), so that they are not pushed to process the word more deeply.

In the same line, Guidi (2009) employed think-aloud protocols to examine the effects of glossing both for meaning and form, as well as type of linguistic item, on reading comprehension, noticing and L2 learning. Following up on Bowles (2004), any mention of a targeted form, lexical or grammatical, accompanied by a pause or a mention of its translation, or a mention of not knowing the meaning of the word was coded as an
instance of noticing. In addition, any comment about the targeted items that indicated an explicit focus on the items as language *per se* (e.g., linguistic analyses of forms and their meanings, connections between meanings and forms made explicit, overt comments about meaning-form relationships, or repetition of a target item more than twice) was coded as an instance of awareness at a higher level than noticing. Sixty-five learners of Spanish were randomly assigned to one of four groups varying in whether they read a text with glosses or without glosses, and whether they were asked to think aloud or not. They completed a reading comprehension test, and production and recognitions tests for both lexical and grammatical items immediately after the treatment and three weeks later. Findings indicated that (a) glossing had a significant effect on text comprehension but not on L2 learning and noticing, (b) overall, noticing significantly correlated with L2 learning, and (c) participants did not engage in language analysis, and did not demonstrate higher levels of awareness of the targeted items in either glossing condition. The researcher reported only a few instances of awareness at a higher level found in both gloss and no-gloss conditions, suggesting that more explicit interventions might be necessary to promote deeper processing and higher levels of awareness.

Finally, Martínez-Fernández (2008a) investigated different levels of awareness in four reading conditions that differed in whether the text included single translation glosses, translation glosses embedded in a multiple-choice task, translation glosses embedded in an fill-in-the-blanks task, or no lexical enhancement. Forty-five English-speaking participants enrolled in college-level second year Spanish language courses completed a pretest, and were randomly assigned to one of four groups. All participants
thought aloud while reading, and completed a retelling task immediately after. Immediate and one-week-delayed posttests measured production and recognition of the targeted words, and the ability to use them in a sentence. In all conditions the targeted words, four concrete and four abstract unfamiliar nouns, occurred four times in the text, but they were bolded and glossed (in the single and multiple-choice gloss conditions) or deleted (in the fill-in-the-blanks condition) only in their first occurrence. The qualitative analysis of the think-aloud protocols in this study brought evidence in support of two different levels of awareness that were found in all conditions: ‘noticing of one word aspect’ and ‘noticing of two word aspects’. ‘Noticing of one word aspect’ includes (a) noticing of meaning only (e.g., verbalizing the meaning of a word, reading the glosses out loud), and (b) noticing of word form only (e.g., verbalizing the word form when commenting lack of knowledge or when trying to infer its meaning unsuccessfully). ‘Noticing of two word aspects’ refers to noticing of both word form and meaning (e.g., verbalizing or writing down the word form and the meaning in the same instance).

This study analyzed differences in vocabulary development and amount of reported awareness between groups, but did not address the correlation between levels of awareness and vocabulary development. The results indicated that the fill-in task led to significantly higher vocabulary development, and involved a significant amount of awareness at the higher level (i.e., noticing of two word aspects). Although the single gloss condition led to awareness at a low level (specifically, noticing of meaning only), vocabulary development in this condition was not significantly different from that of the fill-in condition. Finally, multiple-choice gloss and control conditions resulted in low
vocabulary development and awareness at a low level (specifically, noticing of word form only). Because lack of verbal report does not imply lack of awareness, the researcher admitted that instances of noticing of meaning only in this study might have involved not only reported noticing of meaning but also non-reported noticing of word form. In that case, instances of noticing of meaning only could be interpreted as high awareness. This interpretation might explain why one group showing a significant amount of reported noticing of meaning only (i.e., the single gloss group) was not significantly different in vocabulary development from the group demonstrating higher awareness (i.e., the fill-in condition). Moreover, word frequency might have played a role in noticing the word form in the single gloss group, since participants encountered each targeted word three more times after reporting noticing of meaning only. In any case, findings of this study suggest that future research should revise and refine the operationalization of levels of awareness proposed for lexical items. As Leow et al. (forthcoming) indicate in their review of studies on awareness, a much finer-grained methodological perspective of this construct reveals that awareness should be investigated along the stages of the acquisitional process (construction versus reconstruction), and that variables such as levels of awareness and potential levels of processing at each level of awareness should be carefully considered.

---

9 Leow et al. (forthcoming) distinguish between the stage of construction, which takes place when learners receive and process online the incoming information, and the stage of reconstruction, when learners indicate offline, that is, after they have processed the incoming information, whether they were aware of the targeted item during exposure.
Conclusion

In the last decade, several studies in SLA have investigated the role of awareness in L2 learning employing concurrent verbalizations. None of these studies, however, has investigated the relationship between reported levels of awareness and experiences of recognition memory, but rather the relationship between reported levels of awareness and L2 development. Overall, findings suggest that levels of awareness may impact both the grammatical and vocabulary development that takes place in a variety of learning conditions, including more explicit, less explicit and incidental conditions.

While the existence of different levels of awareness of grammatical items (i.e., noticing and understanding) has been well reported in the literature (e.g., Camblor, 2006; Leow, 1997a; Rosa & O’Neill, 1999; Rosa & Leow, 2004), to date only one study has found evidence in support of different levels of awareness of vocabulary (Martínez-Fernández, 2008a). Bowles (2004) did not find evidence of higher awareness beyond the level of noticing in gloss and no-gloss conditions, and Guidi (2009) found only anecdotal evidence. Therefore, this research area needs to be further investigated.

Moreover, different operationalizations have been proposed for levels of awareness of lexical items. Based on Schmidt’s theoretical framework, Martínez-Fernández (2008a) made a first attempt to operationalize levels of awareness of vocabulary items, and established two levels: noticing of one word aspect (form or meaning) and noticing of two word aspects (both form and meaning). Guidi (2009) proposed an operationalization for levels of awareness of both grammatical and lexical items. According to her operationalization, noticing of one word aspect and noticing of
two word aspects (e.g., any mention of a targeted form accompanied by a pause or a mention of its translation) would constitute only one level of awareness (i.e., noticing), while any comment about the targeted items that indicated an explicit focus on the items as language *per se* would constitute awareness at a higher level than noticing. This operationalization of awareness at a higher level is more in line with the operationalization of ‘understanding’ proposed in previous research on levels of awareness, and clearly reflects a higher level than ‘noticing of two word aspects.’ Future research, therefore, will need to refine previous measures, and establish whether the operationalizations proposed for noticing of one word aspect and noticing of two word aspects reflect different levels of awareness or not, and whether there is evidence for two or three levels of awareness of lexical items (i.e., noticing of one word aspect, noticing of two word aspects, and awareness at a higher level than noticing). The present study aims to address this issue within the input enhancement strand in order to investigate levels of awareness of both grammar and vocabulary in different glossing conditions, as well as their effects on experiences of recognition memory, L2 development and text comprehension.
Input Enhancement

Theoretical underpinnings

Building upon attentional models in SLA that propose a crucial role for attention and noticing in language learning, a large number of empirical studies have investigated different types of instruction and pedagogical techniques that may drive learners’ attention to the linguistic features embedded in the input. Within this framework, an area that has motivated extensive research over the last 15 years is input enhancement. The term ‘input enhancement’ was coined by Sharwood Smith (1991) to define a number of techniques used to make selected features of the input more salient in such a way as to promote noticing and facilitate incidental learning of targeted forms when the main goal is comprehending the input. Furthermore, the researcher distinguished these techniques of ‘external input enhancement’ (e.g., typographical enhancement, input flood, phonological manipulations, etc.) from ‘internal input enhancement’, where particular aspects of the target language become salient at a given stage simply as a result of some natural developmental process outside the learner's control and not because of outside intervention. However, the term ‘input enhancement’ is generally used in the literature to refer to a variety of techniques that involve some kind of external manipulation (i.e., external input enhancement).

Input enhancement techniques are assumed to facilitate incidental learning. The notion of ‘incidental learning’ has been defined as the language development experienced by learners who are processing language for meaning rather than for form (i.e., their goal
is input comprehension rather than grammar or vocabulary learning) and unintentionally learn the targeted items (Schmidt, 1995; Robinson, 2002). According to this definition, learners may or may not pay attention to words and become aware of them while they are processing the input for meaning. Therefore, incidental learning should be distinguished from implicit learning, which takes place outside of awareness. While implicit learning can be incidental only, explicit learning (i.e., learning with awareness) can be intentional or incidental. This view is different from others, where incidental learning is considered to occur when the object of learning is not the focus of attention (cf. Gass, 1999, for a discussion on different views of incidental learning).

In the input enhancement literature, incidental learning has often been operationalized as a learning condition where no instructions to learn the targeted items are provided (cf. Hulstijn, 2001). Empirical studies employing this operationalization usually fail to distinguish external manipulations from internal processes, assuming that learners who do not receive instructions to learn targeted items do not expect to be tested on those items, and thus, do not intend to learn. In order to ensure that incidental learning takes place, it may be necessary to examine learners’ expectations, intentions and processes.

The terms ‘incidental’, ‘intentional’, ‘implicit’, and ‘explicit’ may refer to internal processes but also to external manipulations or conditions that aim to facilitate internal processes. In the context of pedagogical techniques, the terms ‘incidental’ and ‘intentional’, on the one hand, refer to the type of context, more or less communicative, in which the language is processed. In contrast to language-focused tasks, input
enhancement techniques have been defined as incidental and unobtrusive because they allow learners to pay attention to linguistic items embedded in the input within a communicative context. On the other hand, the terms ‘implicit’ and ‘explicit’ refer to the amount of information about the language provided with the input. Input enhancement techniques are usually considered implicit because they present no information about the language. However, the degree of explicitness/implicitness of these techniques may vary depending on the amount of information presented to the learners.

In sum, research on input enhancement is premised on the role of attention and noticing in language learning, and investigates the effects of a variety of techniques that (a) are incidental and unobtrusive, (b) may be more or less implicit, and (c) may lead to both intentional and incidental learning, and both explicit and implicit learning, depending on how the learner interacts with the input.

Two main research strands have investigated different techniques that fall into the category of input enhancement in the written mode, and that are relevant to the present study: (a) research on textual enhancement, which has focused on grammatical (mainly morphosyntactic) development, and (b) research on glossing, which has typically examined the effects of glossed or annotated texts on vocabulary development. A summary of empirical studies on textual enhancement and glossing is presented below.

**Textual enhancement**

Research on textual enhancement has examined the effects of typographical or visual enhancement of the input, such as color-coding, boldfacing, highlighting, or
manipulations of the font and the character size, in written texts. Empirical studies have led to conflicting findings (Shook, 1994; Alanen, 1995; Jourdenais, Ota, Stauffer, Boyson, & Doughy, 1995; White, 1998; Overstreet, 1998; Izumi, 2002; Leow, 1997b, 2001b; Leow, Egi, Nuevo & Tsai, 2003; Bowles, 2003). Jourdenais et al. (1995) investigated the effects of textual enhancement on noticing and written production of Spanish preterit and imperfect forms, and concluded that textual enhancement had a positive effect on both measures. In their study, participants completed a reading comprehension task in one of three conditions: enhancement, non-enhancement, and control. After reading, they were asked to think aloud while writing a description of a series of pictures designed to elicit the target forms. The analysis of the think-aloud protocols revealed that the enhanced group noticed the targeted items more often than the control group, although no evidence for understanding of the aspectual difference between the preterit and the imperfect forms was found in any group. The same pattern was found in the written production posttest, where participants in the enhanced group produced more past tense forms, but did not distinguish accurately between the preterit and the imperfect. The researchers concluded that textual enhancement had a significant impact on noticing and a facilitative role in L2 development.

In contrast, Izumi (2002) found that textual enhancement might have a positive effect on noticing but not on learning. This study provided a treatment over an extended period of time, in which ESL learners were assigned to one of three conditions: output production, input enhancement, and control, during a reading comprehension task. All participants were take notes while reading chunks of a passage. Findings showed that (a)
textual enhancement led to significantly more noticing than the control group, as measured by note-taking, and (b) the output production group significantly outperformed the input enhancement and control groups in both recognition and production of a grammatical targeted form. The researcher concluded that textual enhancement was effective to promote noticing, but not the noticing necessary for learning to take place. The lack of effectiveness of textual enhancement for L2 development has been previously supported by a large number of studies (e.g., Alanen, 1995; White, 1998; Leow, 1997b, 2001b; Overstreet, 1998). Leow (1997b) investigated the effects of textual enhancement and text length on reading comprehension and intake of the Spanish formal imperative, as measured by a multiple-choice recognition test. Findings of this study indicated that text length had a significant effect on reading comprehension while text enhancement did not. Moreover, none of the independent variables appeared to have an impact on intake. Additional evidence for the lack of effectiveness of textual enhancement was found by Overstreet (1998) in a study that addressed the effects of both textual enhancement and content familiarity. Findings of this study showed that none of these variables had an effect on either recognition or production of the preterit/imperfect alternation in Spanish (as measured by a fill-in-the-blank task with multiple choices, and a written narration task). Furthermore, no interaction between content familiarity and textual enhancement was found. In other words, the combination of these factors did not have a positive effect on L2 development, as predicted by the researcher.

Finally, in contrast with Jourdenais et al. (1995) and Izumi (2002), a number of studies that have examined the effects of textual enhancement on both noticing and
learning (Bowles, 2003; Leow, 2001b; Leow, Egi, Nuevo, & Tsai, 2003) have not found a positive effect on either of the dependent measures. Leow (2001b) conducted a study with 74 college students enrolled in beginning Spanish courses. Participants were randomly assigned to enhanced and unenhanced conditions of exposure to the input. Findings showed that there was no significant difference between groups on noticing of the Spanish formal imperative, as measured by think-aloud protocols, and recognition and written production, measured by means of a multiple-choice task and a fill-in-the-blank task, respectively. Noticing, however, was positively correlated with intake (operationalized as recognition) in both enhanced and unenhanced conditions. These findings were further supported by a replication conducted by Bowles (2003) at a higher level of language experience, and by Leow, Egi, Nuevo, & Tsai (2003), who also employed concurrent verbalizations to measure noticing.

Despite the fact that the effect of textual enhancement is premised on the assumption that learners attend to and notice the enhanced forms, very few studies within this strand have attempted to measure noticing (e.g., Jourdenais et al., 1995; Leow, 2001b; Izumi, 2002; Leow et al., 2003; Bowles, 2003). In their review of these studies, Han, Park and Combs (2008) conclude that textual enhancement may induce noticing of the enhanced forms, although noticing does not guarantee acquisition. On the one hand, the researchers argue that most studies that found no effect on noticing (e.g., Leow et al., 2003) have employed single-session treatments, which may not provide enough time for deep processing of the targeted items. In contrast, studies that provide a treatment over an extended period of time do find a positive effect on noticing (e.g., Izumi, 2002).
However, only one study employing multiple treatment sessions has measured noticing; moreover, in this study noticing was measured through note-taking while reading chunks of a passage, a procedure that may provide poor qualitative data compared to think-aloud protocol procedures. On the other hand, the researchers indicate that studies employing single-session treatments should not attempt to measure acquisition, which requires more time to occur. Acquisition certainly requires time and, in order to address it, designs with multiple-session treatments may be preferred over those including only one treatment session, provided that outside exposure is controlled. Yet, acquisition should be distinguished from intake, which may take place after a first encounter with a targeted form, and does not involve system learning. In this sense, several studies employing a single-session treatment have found evidence to support the effect of noticing on intake (e.g., Leow 2001b). Finally, the lack of effects of textual enhancement on L2 development is also supported by Izumi (2002), who administered a treatment in multiple short sessions.

Most studies that find evidence for the efficacy of textual enhancement on L2 development have generally investigated the effects of enhancement in combination with other variables, rather than the effects of simple enhancement (cf. Leow, 2009). In addition, a number of methodological differences across studies may explain the lack of congruence of the findings. In their review of 21 studies on textual enhancement, Han et al. (2008) point out the following methodological differences: “(1) employing simple versus compound enhancement; (2) employing isolated words versus sentences versus discourse as stimuli; (3) enhancing a meaning-bearing versus a non-meaningful form; (4)
employing learners with or without prior knowledge; (5) enhancing the target form many versus one or a few times; (6) using a longer versus a shorter text; (7) employing a single versus multiple short sessions over an extended period of time; (8) enhancing one form versus multiple forms; (9) providing (or not) comprehension support prior to treatment; and (10) providing (or not) explicit instruction on what to focus on prior to the treatment.” (p. 600). Based on the mixed findings of studies reviewed, Han et al. (2008) conclude that the potential efficacy of textual enhancement may depend to some extent on variables such as learners’ prior knowledge, type of linguistic item, and frequency of the enhanced forms.

Another controversial issue in the literature is whether sequential and simultaneous processing of meaning and form may have a differential impact on the results. According to Han et al. (2008), designs promoting sequential processing (i.e., processing meaning before form) may be more efficient in inducing noticing than those requiring simultaneous processing (i.e., processing meaning and form simultaneously). This claim is based on the idea that simultaneous processing of two different types of information that are not automatized can lead to inadequate processing of either or both types of information, and a trade-off effect (e.g., VanPatten, 1990). Robinson (2003), however, has argued in favor of a multiple-resource attentional model that does not support this position. In addition, recent studies (e.g., Leow, Hsieh & Moreno, 2008) have not found evidence for a negative effect of simultaneous attention to form and meaning. Nonetheless, it is important to note that all studies on textual enhancement should measure text comprehension to control for a potential trade-off effect. Moreover,
as Han et al. (2008) claim, reading comprehension should be measured prior to the treatment to control for potential differences on text comprehension. To date, most studies have not implemented this measure.

The scope of research on the effects of textual enhancement has recently been expanded to investigate vocabulary development. Kim (2006) investigated whether lexical elaboration (explicit, implicit or unelaborated), typographical enhancement (enhanced versus unenhanced), or a combination of both affected learning of English vocabulary in a sample of 297 Korean learners of English. The results of this study showed that typographical enhancement did not aid form and meaning recognition of vocabulary, and that both explicit and implicit lexical elaboration (i.e., providing the meaning of a word versus providing an apposition within the text, respectively) aided meaning recognition of vocabulary, regardless of whether the text was further enhanced or not. These findings add to previous studies supporting a lack of effect of simple textual enhancement on morphosyntactic development, but contrast somehow with studies on glossing which find a beneficial effect for glossing (i.e., explicit lexical elaboration) on vocabulary recognition but not for appositives (e.g., Watanabe, 1997). Findings of empirical studies on glossing are reviewed in the next section.

Glossing

A gloss is defined as a definition, synonym or translation that is provided for a particular obscure word at the bottom or margin of a text. Glossing is a technique that has been traditionally used to aid learners when they read an L2 text and encounter unknown
words. When encountering unfamiliar words in non-manipulated texts, some learners may try to infer their meaning in order to understand the text. However, the meaning of some words is difficult to infer, the context does not always facilitate inferences, and some learners are not good at inferring. Moreover, many readers often skip unknown words while reading. For this reason, teachers and materials designers have used glossing as an effective technique to aid L2 reading comprehension.

Although the main purpose for glossing was initially to aid reading comprehension, researchers have started to consider glossing as a technique of input enhancement that may affect vocabulary development. On the one hand, glossed texts provide learners with a rich input where lexical items are highly contextualized, and on the other hand, glossing enhances lexical items without interrupting the reading process. Therefore, among the different input enhancement procedures, glossing may be considered an incidental and unobtrusive technique (i.e., targeted items are processed in a communicative context), that is more explicit than simple textual enhancement since it provides the meaning of the targeted items (i.e., information about the language), and that can show different degrees of explicitness depending on how much information the glosses provide (e.g., a translation or synonym only, a combination of a definition and an example, a combination of a translation and an image, a combination of lexical and grammatical information, etc.).

In the last 20 years, a growing number of empirical studies have tested the effectiveness of glosses on both text comprehension and vocabulary learning (e.g., Hulstijn, 1992; Watanabe, 1997; Bowles, 2004; Guidi, 2009). Research studies have
examined different aspects of glosses, including (a) L1 translation glosses versus L2 synonym or definition glosses, and students’ preferences on L1 or L2 glosses (Jacobs, Dufon, & Hong, 1994; Bell & LeBlanc, 2000; Hee Ko, 2005); (b) glosses for frequent words (words that occur more than once in the text) versus glosses for non-frequent words (Hultijn, Hollander & Greidanus, 1996); (c) computerized versus paper-and-pen glosses (Bowles, 2004), and students’ attitudes toward both types (Davis & Lyman Hager, 1997; Roby, 1999); (d) text glosses versus multimedia annotations, including pictorial, audio or video annotations (Chun & Plass, 1996; Plass, Chun, Mayer, & Leutner, 1998; Kost, Foss, & Lenzini, 1999; Al-Seghayer, 2001); (e) strategies induced by different types of glosses (Rott, 2005; Hee Ko, 2005); and (f) students’ look up behavior (Bell & LeBlanc, 2000; Hulstijn, 1993; Lomicka, 1998; Plass, Chun, Mayer, & Leutner, 1998). Additionally, glossing has been compared to other non-computerized procedures to access words’ meanings while reading, such as dictionary use, concise context, meanings embedded in appositions, glosses embedded in multiple-choice tasks, and glosses embedded in fill-in-the-blank tasks, among others (e.g., Hulstijn, 1992; Hulstijn, Hollander & Greidanus, 1996; Watanabe, 1997; Hulstijn & Laufer, 2001; Rott, 2005; Rott, Williams, & Cameron, 2002; Martínez-Fernández, 2008a). Studies comparing different procedures have been concerned mainly with vocabulary development, although some of them investigate also their effectiveness in text comprehension. The next sections present a summary of empirical studies that examine (a) effects of glosses on text comprehension, (b) effects of glosses on vocabulary development, and (c) effects of
type of glossing and other lexical procedures on vocabulary development and text comprehension.

**Effects of glosses on text comprehension**

Early studies that have investigated the effects of glosses on reading comprehension do not support the effectiveness of glosses (e.g., Johnson, 1982; Pak, 1986; Jacobs, Dufon, & Hong, 1994). Johnson (1982) found that ESL learners who read a glossed text with definitions in the margin did not outperform those who read a non-glossed text on a recall protocol written in the L2, and a 3-week delayed cloze test. The researcher concluded that glosses might preclude readers from global comprehension, leading them to focus on word-by-word decoding, and thus approach the task from a bottom-up processing perspective.

This conclusion was also supported by Pak (1986). In this study, ESL learners with high and low reading ability were randomly assigned to one of three conditions varying in whether they read (a) a text with word definitions in the margin, (b) a text with word definitions embedded in the passage, and (c) a non-glossed text. Findings indicated that reading ability had an effect on text comprehension, as measured by a cloze test, but the experimental groups did not outperform the control group.

Finally, Jacobs, Dufon, and Hong (1994) investigated the effects of L1 and L2 glosses. Participants in this study were 85 intermediate learners of Spanish, who were assigned to L1 gloss (English), L2 gloss (Spanish), and no-gloss conditions. The researchers found no effect for gloss conditions on text comprehension, as measured by a
recall protocol written in the L1. However, participants with higher proficiency level seemed to benefit more from glosses than participants with lower proficiency level. In addition, the analysis of students’ preferences revealed that participants preferred glossed to non-glossed texts, and L2 glosses to L1 glosses. The researchers concluded that the use of glosses might be beneficial but only for those learners who reach a certain level of proficiency.

In contrast to the findings previously reported, several studies have found evidence to support the effectiveness of glosses in reading comprehension (Davis, 1989; Jacob, 1994; Hee Ko, 2005; Bell and Le Blanc, 2000; Bowles, 2004; Guidi, 2009). Davis (1989) investigated whether glosses improved the comprehension of an L2 literary text. Participants were 71 intermediate learners of French, randomly assigned to one of three conditions varying in type of exposure: (a) text with questions and comments to guide the reader, and definitions of vocabulary items, presented before reading, (b) text with the same questions and vocabulary definitions, but in glossed format, and (c) text with no aids (control). Findings showed that the experimental groups outperformed the control group on a recall protocol written in the L1. In this study, however, the effect of glosses was not isolated from the effect of the guiding questions and comments. The researcher concluded that glosses, rather than distracting readers, ensured more fluent reading of the passage. Jacobs (1994) did isolate the effect of glosses in a study with 166 fourth-semester college students of Spanish, who were split into two conditions: gloss (English translation in the margin) and no-gloss. The results showed that the gloss group significantly outperformed the no-gloss group on a recall protocol written in the L1.
Recent studies have used other text comprehension measures different from cloze tests and written recall tasks, such as multiple-choice tests (e.g., Hee Ko, 2005; Bell and Le Blanc, 2000; Bowles, 2004; Guidi, 2009). Hee Ko (2005) investigated the effects of L1 and L2 glosses on text comprehension, students’ preferences, and strategies induced by glosses, in two different studies. In a quantitative study, 94 intermediate learners of English split into three conditions: L1 gloss (Korean synonym or definition), L2 gloss (English equivalent or definition), and no gloss. In a qualitative study, 12 participants that did not participate in the quantitative study were assigned to the same three conditions, and were asked to think aloud while reading. The results showed that only the L2 gloss group significantly outperformed the control group on a multiple-choice comprehension test. However, the reliability of this test was low. With respect to reading strategies, as measured by think-aloud protocols, the gloss groups used more high-level strategies, such as making inferences, while the control group used more low-level strategies, such as guessing and skipping difficult words. Based on the think-aloud protocols, the researcher concluded that both L1 and L2 glosses facilitated reading comprehension. Finally, the analysis of students’ preferences showed that most participants preferred L2 glosses to L1 glosses if they could understand them.

In a computerized study, Bell and LeBlanc (2000) also investigated the effects of L1 and L2 glosses on reading comprehension. Forty fourth-semester learners of Spanish were assigned to one of two conditions: L1 gloss (English) or L2 (Spanish) gloss. Participants read a text containing 67 words linked to a gloss, and answered 10 multiple-choice comprehension questions. Contrary to the findings reported by Hee Ko (2005) on
traditional glosses, in this study it was found that both types of gloss correlated with higher performance on the comprehension test, suggesting that the language used in glosses is not a crucial factor for comprehension.

Guidi (2009) examined the effect of computerized L1 glosses on text comprehension as measured by a multiple-choice questionnaire. Sixty-five learners of Spanish were randomly assigned to one of four groups varying in (a) whether they read a glossed text or a non-glossed text, and (b) whether they were asked to think aloud or not. Findings indicated that the gloss groups significantly outperformed the non-gloss groups in text comprehension. Additionally, in contrast to Hee Ko’s (2005) findings, the analysis of the think-aloud protocols in this study revealed that participants in both gloss and non-gloss conditions read the text in a bottom-up fashion, without making inferences, activating prior knowledge, making connections between paragraphs or recapitulating ideas. Thus, this study supported the effectiveness of glosses for reading comprehension at a micro level. Furthermore, the researcher concluded that glossing did not have a detrimental effect by precluding participants from global comprehension, and leading them to focus on word-by-word decoding, as Johnson (1982) had argued, since the same reading strategies were found in both the gloss and no-gloss groups.

The increasing use of computerized glosses led Bowles (2004) to investigate the effects of paper-and-pen versus computerized L1 glosses. Fifty beginning learners of Spanish split into three groups: computer, non-computer and control conditions. Participants thought aloud while reading the text, and answered 12 multiple-choice comprehension questions in English immediately after reading. They could refer back to
an unglossed version of the text in order to answer the questions. Results showed that gloss conditions performed significantly higher than the control group, and that there was no significant difference between the computer and no-computer groups.

While the computer does not seem to impact the effectiveness of short definitions provided for target items in the margin of a text, the use of technology has allowed researchers to expand both the typology of glosses and the possibilities of research designs, in which experimental texts containing words hyperlinked to one or more types of glosses are employed. Recently, a number of studies have investigated the effects of simple textual glosses versus multimedia annotations (e.g., Lomicka, 1998; Davis & Lyman-Hager, 1997; Chun & Plass, 1996; Plass, Chun, Mayer, & Leutner; 1998). The assumption in these studies is that multimedia annotations may impact reading comprehension both at the micro level (i.e., word decoding, local comprehension) and macro level (i.e., global comprehension). However, overall findings of these studies are not conclusive regarding the effectiveness of multimedia annotations, and do not find support for a positive effect of traditional vocabulary glosses on text comprehension.

To investigate the effects of multimedia annotations, Lomicka (1998) conducted a pilot study with twelve second-semester learners of French, who were assigned to one of three conditions varying in the type of information they could access: simple gloss (L2 definition and L1 translation), multimedia gloss (L2 definition, L1 translation, image, reference, question, and pronunciation), and control conditions. Participants read some parts of a poem, and were asked to think aloud and verbalize their understanding of the text, the reason why they chose to access each gloss, and any other thought they might
have while reading. The number of inferences (i.e., explanations, paraphrases, associations, predictions) coded in the think-aloud protocols was taken as a measure of reading comprehension. Only data regarding the number of explanations generated while thinking aloud were submitted to a statistical analysis. Findings indicated that there was no difference among groups in amount of explanations. Additionally, the analysis of the look-up behavior registered by a tracking mechanism revealed that participants preferred to access L1 translation glosses even if they had the option to access other types of glosses. Despite these results, the researcher concluded that multimedia annotations aided comprehension more than simple textual glosses, and that participants should receive instruction to take advantage of the benefits of multimedia annotations. Although this study clearly indicates a lack of effect of both textual and multimedia annotations, this result can be questioned because the measure used for text comprehension did not measure comprehension but rather the reading process.

Similar findings were reported by Davis and Lyman-Hager (1997) in a study with English-speaking intermediate learners of French. In this study, participants read a computerized text, and were given the option to access L1 and L2 definitions, grammatical explanations, aids for intratextual reference relations, pictures and cultural notes. The results revealed that participants looked up L1 translation glosses more often than any other type of gloss. Moreover, the aid of L1 translation glosses was not correlated with text comprehension, as measured by a written recall protocol and a multiple-choice test. However, the overall difficulty of the text might explain this result,
according to the researchers. In addition, they concluded that learners needed to be trained in order to use multimedia annotations and reach higher text comprehension.

Finally, Chun and Plass (1996) conducted three computerized studies with second year college learners of German, who were asked to read a text where targeted items were annotated in three different conditions: (a) traditional textual gloss, (b) traditional textual gloss plus visual annotation (i.e., a picture or a video), and (c) no gloss. In addition, all participants watched a video preview of the content of the passage. This pre-reading activity was included to provide participants with some prior knowledge that they could activate while reading. Findings showed that a combination of textual and visual annotations led to the highest performance on text comprehension, as measured by a written recall protocol, while textual glosses led to the lowest performance. Moreover, the recall protocols revealed that participants mentioned ideas that were included in the video preview to a greater extent than ideas that were not included. The researchers concluded that (a) multimedia annotations that provide a definition and a visual stimulus had a positive impact on text comprehension at a micro level, thus strengthening bottom-up processes, and (b) pre-reading activities such as the video preview might activate top-down processes of reading comprehension (i.e., prior knowledge activation), thus affecting text comprehension at a macro level. It is not clear, however, why the text with traditional glosses only led to lower performance than the control group.

In a follow-up study, Chun, Mayer, and Leutner (1998) suggested that the effect of textual glosses versus visual annotations might depend on learners’ preferences for one type or the other. This study investigated the effects of different types of glosses
according to the learners’ preferred mode: textual, visual or both textual and visual. One hundred three second year college learners of German read a computerized text at their own pace, consulting the type of annotations that they preferred. One day after, they were asked to write everything they could remember of the passage. Participants were grouped according to their preference for glosses using log-file data from the computer program used. They were divided into verbalizers, visualizers, and those having no strong preference. The results showed that participants recalled the text better when they could choose the gloss in their preferred mode. Therefore, this study does not support an effect or a lack of effect for either textual or multimedia annotations, suggesting that their effectiveness may interact with other factors such as learners’ preferences.

To summarize, previous research on the effects of glosses on reading comprehension has yielded mixed results. Some studies found a lack of effect (Johnson, 1982; Pak, 1986; Jacobs, Dufon, & Hong, 1994; Lomicka, 1998; Davis & Lyman-Hager, 1997; Chun & Plass, 1996), while others supported a positive effect (Davis, 1989; Jacob, 1994; Hee Ko, 2005; Bell & Le Blanc, 2000; Bowles, 2004; Guidi, 2009). The different measures used for text comprehension across studies may explain to some extent these conflicting findings. Measures used in the research studies reviewed include (a) cloze tests, (b) written recall protocols, (c) think-aloud protocols, and (d) multiple-choice questionnaires. Cloze tests present a text where some words have been deleted and require learners to fill in the blanks. This is a controversial measure of text comprehension because it is not clear whether this test measures text comprehension, global reading skills, vocabulary and grammar knowledge, or other abilities or
knowledge. For this reason, this measure was employed in early studies (e.g., Johnson, 1982; Pak, 1986) but it has not been used later. Regarding written recall protocols, the most important limitation of this measure is that memory may play a role when completing this task since participants may understand the content of a passage and not be able to remember all of it. Therefore, written recall protocols measure both comprehension and memory. Moreover, recall protocols written in the L2 (e.g., Johnson, 1982) measure also vocabulary and grammar knowledge and L2 writing skills. Finally, think-aloud protocols (e.g., Lomicka, 1998) seem more appropriate to measure reading processes and strategies rather than text comprehension. For these reasons, other types of tests, such as multiple-choice and open-ended questionnaires, may be more appropriate measures of text comprehension.

Overall, studies showing positive effects for traditional glosses operationalized as short definitions in the L1 or the L2, in computerized or paper-and-pen formats, have employed more robust measures of text comprehension (e.g., Bell and Le Blanc, 2000; Bowles, 2004; Guidi, 2009) than studies supporting a lack of effect. Thus, there seems to be some robust evidence to support the effectiveness of glosses as a pedagogical technique to aid text comprehension.

**Effects of glosses on vocabulary learning**

Some of the previous studies reviewed have examined not only the effectiveness of glosses in reading comprehension but also their effects on vocabulary learning (Jacobs, Dufon, & Hong, 1994; Bowles, 2004; Guidi, 2009; Chun & Plass, 1996; Plass et al.,
1998). Overall, these studies show that gloss groups perform significantly higher than control groups on vocabulary tests, although the gain seems to be generally low. Jacobs, Dufon, and Hong (1994) investigated the effects of L1 and L2 glosses. Participants in this study were 85 intermediate learners of Spanish, who were assigned to L1 gloss (English), L2 gloss (Spanish), and no-gloss conditions. Vocabulary learning was measured by a translation test in which participants were presented with a list of the targeted words, and were asked to provide the L1 equivalent. Findings on vocabulary learning indicated that both L1 and L2 gloss conditions outperformed the control group but this effect was not retained four weeks after.

In her study on the effects of paper-and-pen versus computerized L1 glosses, Bowles (2004) examined the effectiveness of glossing in both vocabulary recognition and production, as measured by a multiple-choice test and a translation test, respectively. In the production test, participants (fifty English-speaking learners of Spanish) were presented with a list of 40 words in English (21 targeted items and 19 distractors) and were asked to provide their Spanish equivalent. The multiple-choice test included 21 target items, for which four possible translations in English were provided. Participants completed the recognition test after the production test so that the multiple-choice questions would not provide participants with additional exposure to the targeted items. The results of immediate and 3-week delayed posttests showed that the gloss groups significantly outperformed the control group in all measures, and performance on the recognition test was higher than on the production test.
Guidi (2009) investigated the effectiveness of glossing on recognition and production of vocabulary and different grammatical items. In this study, production was measured by a fill-in-the-blank test including 30 sentences. Participants (sixty-five English-speaking learners of Spanish) were asked to fill in the blank in each sentence by providing the Spanish equivalent of the English word or phrase that appeared in parenthesis. Recognition was measured by a multiple-choice test that included 30 target items, for which three possible translations in English and a ‘none of the above’ option were provided. Findings of this study did not provide enough evidence to support a claim for a glossing effect. However, the results supported a combined effect of gloss with type of linguistic item. The researcher concluded that the effect of glosses on L2 learning might interact with factors relative to the inherent characteristics of different types of items. In addition, this study suggests that other factors, such as frequency of exposure and number of targeted items embedded in the input, may interact with a glossing effect.

Finally, findings of studies comparing the effects of textual glosses, pictorial glosses and/or dynamic visual glosses on vocabulary learning are not conclusive. Kost, Foss, and Lenzini (1991) investigated the effects of textual and pictorial glosses. Fifty-six second-semester learners of German were assigned to one of three conditions: textual, pictorial, and a combination of the two. Vocabulary learning was measured by a production test and two recognition tests administered right after reading the text, and two weeks later. Results showed that the group accessing combined glosses significantly outperformed the other two groups in the recognition tests but not in the production tests.
Chun and Plass (1996) investigated the effects of textual glosses, a combination of textual and pictorial glosses, and a combination of textual and dynamic visual glosses in a study with a within-subjects design. Findings indicated that participants (160 second-year learners of German) performed much better on vocabulary recognition tests than on production tests for items in all conditions, and that scores for items annotated with text and picture were significantly higher than those for items glossed with text only or with text and video.

Al-Seghayer (2001) further investigated the effect of dynamic video and still picture conditions in a within-subject design study. Thirty learners of English as a second language read a text that included words glossed with text, text along with still picture, and text along with dynamic video. In contrast to findings reported by Chun and Plass (1996), results of recognition and production posttests in this study indicated that learners’ performance was higher for words annotated with text and video than for words annotated with text and picture, or with text only.

Following up on previous studies, Plass et al. (1998) compared the effects of textual glosses, visual glosses and a combination of textual and visual glosses in relation to learners’ preferences. The researchers found that participants performed better on the vocabulary posttest, completed one day after the treatment, when both visual and textual information were selected, moderate when only one mode was selected, and worse when neither was selected.

To conclude, this survey of previous research shows that glossing a text may be beneficial for vocabulary learning, especially for vocabulary recognition, in comparison
with non-glossed texts (e.g., Jacobs, Dufon, and Hong, 1994; Bowles, 2004). However, the effectiveness of glossing may depend on a variety of factors, such as type of linguistic item and frequency (e.g., Guidi, 2009). When traditional glosses (i.e., short L1 or L2 definitions) are compared to visual glosses or glosses combining textual and visual stimuli, findings suggest that a combination of textual and visual annotations has a higher impact on vocabulary development than textual glosses (e.g., Kost, Foss, and Lenzini, 1991; Chun and Plass, 1996; Plass et al., 1998; Al-Seghayer, 2001). Overall, studies comparing textual and visual glosses do not include control conditions. Therefore, they do not address the effect of textual glosses on vocabulary learning, but only their effect in relation to other types of glosses.

Within the glossing strand, researchers have become interested in exploring the effectiveness of glossing in combination with different types of lexical interventions that are more or less incidental or language-focused, and that may vary in their degree of explicitness/implicitness. Several studies have compared the effects of traditional glosses and other procedures, such as multiple-choice glosses, dictionary use, appositions, concise context, and fill-in-the-blank tasks on L2 vocabulary learning (e.g., Hulstijn, 1992; Hultijn, Hollander & Greidanus, 1996; Watanabe, 1997; Nagata, 1999; Hulstijn & Laufer, 2001; Rott, Williams, & Cameron, 2002; Williams & Rott, 2003; Rott, 2005; Kim, 2008; Keating, 2008; Martínez-Fernández, 2008a, 2008b, 2009b). A summary of these studies is presented in the next section.
Effects of type of glossing and other lexical procedures on vocabulary learning and text comprehension

Most research studies have shown that glossing has a positive effect on text comprehension, as well as on vocabulary learning, although the vocabulary gain is generally very low (e.g., Bowles, 2004; Guidi, 2009). A number of researchers have explained this low vocabulary gain by arguing that glosses may preclude learners from making a mental effort (e.g., Hulstijn, 1992; Hulstijn et al., 1996; Watanabe, 1997). According to them, making a mental effort to get the meaning of a word (e.g., engaging in inference or search processes) may have a greater impact on vocabulary learning than accessing the meaning without a mental effort. In order to test this prediction, researchers have conducted numerous investigations that compare the effects of glosses and different procedures that require a mental effort on the learners’ part while guiding them to get words’ meanings.

The assumption that making a mental effort had a positive effect on vocabulary learning led Laufer and Hulstijn (2001) to propose the Involvement Load Hypothesis within the incidental L2 vocabulary learning strand. This hypothesis is rooted in the levels of processing framework proposed in cognitive psychology, on the one hand, and the attentional models in SLA, on the other. In line with cognitive psychology research, Hulstijn (2001) rejects the initial view of levels of processing (i.e., semantic versus perceptual processing), and understands depth of processing as elaboration and amount of attention. According to the researcher, “processing new lexical information more elaborately (e.g., by paying attention to the word’s pronunciation, orthography,
grammatical category, meaning and semantic relations to other words) will lead to higher retention than by processing new lexical information less elaborately (e.g., by paying attention to only one or two of these dimensions)” (p. 270). Laufer and Hulstijn (2001) proposed the notion of ‘involvement’ as an operationalization for the construct of depth of processing in SLA.

The Involvement Load Hypothesis posits that incidental tasks inducing higher involvement are conducive to the type of processing that is deemed crucial for vocabulary retention. The notion of involvement includes three task-specific components: a motivational component, ‘need’ (+N), and two cognitive components, ‘search’ (+S) and ‘evaluation’ (+E). ‘Need’ is defined as “the drive to comply with task requirements, whereby the task requirements can be either externally imposed (i.e., moderate need, +N) or self-imposed (i.e., strong need, ++N)” (p. 14). ‘Search’ and ‘evaluation’ require allocating attention to form-meaning relationships. The former is defined as the attempt to find the meaning of an unknown word, and the latter involves “a comparison of a given word with other words, a comparison of a specific meaning of a word with its other meanings, or combining the word with others in order to assess whether a word (i.e. a form-meaning pair) does or does not fit its context” (p. 14). In addition, the researchers also distinguished ‘moderate evaluation’ (+E), when words being evaluated must fit in a given context, from ‘strong evaluation’ (++E), when words being evaluated must be combined with additional words in an original context, that is, a context created by the learner. Depending on the presence or absence of these three components, tasks may induce different degrees of involvement on the learner’s part that may lead to noticing
and elaborated processing of the words, and that will affect vocabulary retention as a result. Finally, the researchers suggest that (a) task effectiveness does not depend on whether the task is input or output oriented, but only on its involvement load, and (b) the effect of involvement load may interact with other factors such as type of item and quantity of exposure.

Based on these theoretical underpinnings, several empirical studies have tested the effectiveness of the following procedures and tasks, designed to facilitate vocabulary learning by inducing a mental effort without interrupting the reading process: (a) appositives, which provide the definition of the targeted items within the text (Watanabe, 1997); (b) concise context, which is assumed to facilitate learners’ inferences (Hulstijn, 1992; Watanabe, 1997); (c) periodic L2 text reconstruction with opportunities to recheck input (Rott, Williams, & Cameron, 2002; Rott & Williams, 2003); (d) gloss-retrieval, which involves glossing of the targeted word during the first encounter and answering a question about its meaning in a subsequent encounter (Rott, 2007); (e) multiple-choice glosses, which display several translations from which the reader chooses (Hulstijn, 1992, Watanabe, 1997; Rott, Williams, & Cameron, 2002; Rott & Williams, 2003; Rott, 2005; Martínez-Fernández, 2008a); and (f) fill-in tasks, which display several glossed words from which the reader chooses the appropriate word to fill in the blanks of a text (Hustijn & Laufer, 2001; Keating, 2008; Kim, 2008; Martinez-Fernández, 2008a, 2008b, 2009b). In addition, the effectiveness of glosses has been compared to that of dictionary use while reading (Hulstijn et al., 1996). A summary of these studies is presented below.
Hulstijn et al. (1996) investigated the effects of L1 marginal glosses, access to a bilingual dictionary, and control conditions, as well as word frequency as a within-subject variable. Participants were 78 Dutch learners of French, split into the three conditions. They were given 25 minutes to read a fictional narrative text, and were not told that they would be tested on the targeted items. Following the treatment, they completed a text comprehension questionnaire, and immediate vocabulary posttests. Findings showed that (a) the gloss group significantly outperformed the other groups on two translation tests in which target words were given in isolation and in context respectively, and (b) participants benefited from word frequency. Yet, the means of the gloss group for non-frequent and frequent words in these posttests were very low (i.e., 1.6 and 3.4 out of 16 targeted words, respectively). The fact that participants in the dictionary condition seldom used their dictionary explains the lack of effectiveness for this condition. However, the results indicated that when participants did look up a word (look-up behavior was controlled by a post-exposure questionnaire), their chance of remembering its meaning was greater than the average retention in the gloss group. The researchers suggested that meaning search might lead to deep processing, while meaning provision would not lead to deeply process the targeted words. In order to engage in meaning search, however, the targeted words needed to be perceived as relevant in the context of the reading goal.

Hulstijn (1992) conducted five experiments to investigate whether retention of inferred meanings was higher than the retention of given word meanings. Two experiments compared three meaning-to-be-inferred procedures (multiple choice gloss,
concise context, and no-cue or control) and one meaning-given procedure (L1 translation/ L2 synonym gloss), and three experiments compared only two procedures: multiple-choice glosses and single glosses (L2 synonym gloss). Findings of these experiments are difficult to compare due to a number of methodological differences across them: learners of Dutch versus native speakers, L2 words versus pseudo-words, between-subject design versus within-subject design, different prior knowledge and retention measures, intentional versus incidental learning conditions, inclusion of a control group versus lack of control group, and different number of choices given in the multiple-choice procedure (from 2 to 4).

Experiments I, III, and V were conducted with L2 learners but only Experiment I included a control group. In this experiment, 65 learners of Dutch split into four conditions: concise context, L1 single gloss (Turkish translation), L1 gloss embedded in a multiple-choice task with four choices, and control. In the concise context condition, a sample sentence was given in the margin, providing a concise context from which the meaning of the targeted item could be easily inferred. All participants were asked to answer comprehension questions while reading to ensure that they were processing the text for its content, and completed two unannounced vocabulary tests immediately after. Vocabulary measures included a translation posttest where participants had to provide L1 equivalents for the targeted words, and a production cloze test, which required them to produce the targeted L2 words. The results showed that the single gloss group significantly outperformed concise context, multiple-choice gloss, and control conditions.
on the immediate translation posttest, thus running counter to the researcher’s predictions. No difference between groups was found in the production cloze test.

Experiment III investigated the effects of L2 single glosses and L2 glosses embedded in a multiple-choice task with four choices. Participants were 45 learners of Dutch, and all of them read a text containing 6 targeted items with L2 single glosses and 6 targeted items with multiple-choice glosses. Vocabulary learning was measured by a test in which the target words appeared in the original context, and participants were asked to provide the L1 equivalent. Findings of this experiment showed that the multiple-choice group outperformed the single gloss group. However, the vocabulary test employed is not a valid measure of vocabulary learning since providing the original context of the targeted words might have helped learners infer their meaning during test completion. This potential test effect was not controlled since there was no control condition.

Finally, Experiment V compared L2 single glosses and L2 multiple-choice glosses under intentional and incidental conditions of exposure to the input. Participants were 35 learners of Dutch who split into two conditions: single gloss versus multiple-choice gloss. They first completed the incidental treatment, where vocabulary posttests were not announced. After completing the vocabulary posttests, they were administered the intentional treatment, followed by vocabulary posttests. Vocabulary learning was measured by two meaning tests in which they were presented with the targeted items in isolation and in context, respectively, and had to provide the L1 equivalents. Findings showed that (a) the intentional condition led to significantly higher vocabulary learning
than the incidental condition, and (b) although the multiple-choices glosses led to higher vocabulary learning, there was no significant difference between single gloss and multiple-choice gloss conditions.

The researcher concluded that the multiple-choice procedure seemed to have a higher retention effect, although he also suggested that it should only be employed in a classroom context where a teacher is present, given that many participants in all experiments often selected the wrong meanings. Finally, findings regarding comprehension tests were not reported or discussed.

In order to minimize the possibility that learners selected the wrong meanings, Watanabe (1997) investigated the effects of a multiple-choice gloss task in which only two alternatives instead of four were displayed. Participants in this study were 231 Japanese learners of English. Half of them read a text under one of five conditions of exposure: L2 single gloss, L2 multiple-choice gloss, appositive (i.e., definition of the targeted items in embedded appositive clauses within the text), non-glossed original text, and control. The other half split into the same five conditions, but was asked to write down a translation of the targeted words as they read. All participants completed a meaning production pretest immediately before the treatment, answered five open-ended comprehension questions after reading the text, and completed an unexpected vocabulary immediate posttest (identical to the pretest), and two one-week-delayed posttests in which participants had to provide the meanings of the targeted words in isolation and in its original context. The results revealed that (a) the translation task did not have an effect on vocabulary learning, (b) both the single gloss and the multiple-choice gloss groups
significantly outperformed the appositive and control groups on the vocabulary posttests; and (c) the loss from the immediate to the delayed vocabulary posttests was significantly higher for the appositive and control groups. However, the multiple-choice gloss condition did not outperform the single gloss condition. The researcher concluded that students in these two groups did establish form-meaning connections, while students in the other groups did not pay attention to the targeted words, or did not notice or establish form-meaning connections between the words and the appositives. However, attention, noticing or form-meaning connections were not measured in this study. Finally, findings on text comprehension revealed that only the single gloss group performed significantly better than the control group.

Following up on previous studies, Nagata (1999) investigated the effectiveness of computerized L1 multiple-choice glosses that displayed only two choices, and where learners’ selections were followed by immediate feedback (i.e., a right/wrong indication). In this study, 26 learners of Japanese took a pretest two days before the experiment, and were assigned to one of two groups: an L1 single gloss group and an L1 multiple-choice gloss group. Targeted items included 20 lexical items that appeared with different frequency in the text, and 6 grammatical items related to three grammatical structures. Participants read the text, and completed translation immediate and one-month-delayed posttests. Results indicated that the multiple-choice gloss group significantly outperformed the single gloss group on the lexical and grammatical immediate posttests and on the grammatical delayed posttest. The researcher concluded that participants in the multiple-choice condition paid more attention to the targeted items, and made greater
mental efforts to interpret them. However, besides the fact that attention or mental effort were not measured, this study moves away from the incidental learning strand by introducing the role of feedback. In addition, the study did not control for a potential trade-off effect on text comprehension.

Rott, Williams and Cameron (2002), Rott and Williams (2003), and Rott (2005) investigated the effects of multiple-choice glosses within a form-meaning connections framework (cf. VanPatten, Williams, Rott, and Overstreet, 2004 for a discussion on form-meaning connections). In a study with seventy-six fourth-semester learners of German, Rott, Williams and Cameron (2002) assigned intact classes to one of four conditions of exposure to the input: L1 multiple-choice glosses, periodic L2 text reconstruction with opportunities to recheck input, combined treatments, and control. Participants completed a translation pretest, and were allotted either 30 minutes (in the conditions with L2 text reconstruction) or 20 minutes (in all other conditions) to read the text. Each targeted word (4 concrete nouns) occurred four times. In the gloss condition, glosses with four options (including a “don’t know” option) were provided in the first and fourth encounters. The results showed that the multiple-choice gloss condition significantly outperformed the other conditions on immediate production and recognition vocabulary posttests, and only the combined treatment condition retained vocabulary recognition gains five weeks later. In addition, only the combined treatment condition significantly outperformed the control group on text comprehension as measured by a written recall task. The researchers concluded that multiple-choice glosses had a positive impact on vocabulary learning, but
did not lead to the level of deep processing that is claimed to promote long-term retention of form-meaning connections.

A qualitative follow-up investigation conducted by Rott and Williams (2003) explored whether learners in the multiple-choice condition approached the task in the way that the researchers expected. In this study, 12 fifth-semester learners of German where randomly assigned to one of two conditions: a no-gloss condition involving L2 text reconstruction only, and a gloss condition involving a combined treatment of L2 text reconstruction and L1 multiple-choice glosses. Participants read the text silently first, and then read it again engaging in a think-aloud procedure. The think-aloud protocols were analyzed in order to find out what strategies learners used at each encounter with the targeted word: use of semantic context and gloss, use of grammatical information, no elaboration, L2 verbalization of the targeted word, and ignoring. The first two strategies indicated that learners made form-meaning connections, while the others did not.

Noticing was measured by a post-exposure questionnaire, in which participants were presented with the targeted words and indicated whether they remembered having seen them in the text or not. Findings showed that both groups noticed all targeted words. Scores on immediate production and recognition posttests showed that the gloss condition outperformed the no-gloss condition, and the think-aloud protocols revealed that multiple-choice glosses triggered search and evaluation processes. The researchers concluded that glosses helped establish form-meaning connections, and that L2 text reconstruction in the gloss condition might have strengthened those connections. The
effect of multiple-choice glosses, however, was not isolated from the effect of L2 text reconstruction in this study.

To further investigate the processing behavior induced by multiple-choice glosses, Rott (2005) conducted a study with 10 third-semester learners of German. Participants completed a pretest one week prior to the treatment, and were randomly assigned to one of two conditions of exposure: L1 single gloss and L1 multiple-choice gloss conditions. All participants read the same text employed by Rott, Williams and Cameron (2002) and Rott and Williams (2003). However, in this study glosses were provided for the targeted words only in their first occurrence. Participants first read silently a version without glosses, and then were asked to read the version with glosses and think aloud while reading. Text comprehension was measured by an L1 written recall task. Immediate and four-week delayed vocabulary measured the ability to recognize the word form, provide an L1 equivalent, use the word in an L2 sentence, and recognize its meaning from multiple choice options. Findings revealed that (a) both groups performed equally well on the immediate vocabulary posttests, but the multiple-choice condition showed retention of more words and more word aspects four weeks later; and (b) both groups showed low comprehension, and comprehended the main ideas at the same level, but the multiple-choice group recalled more supporting ideas.

The analysis of the think-aloud protocols indicated that the multiple-choice group integrated both meta-cognitive strategies (e.g., monitoring word comprehension, verbalizing the targeted words, and referring to the gloss —noticing) and semantic-elaborative strategies (e.g., using background knowledge and context, and hypothesis-
testing) to establish form-meaning connections during the first encounter with a new word. In contrast, the single gloss group used only meta-cognitive strategies. Both groups noticed most targeted words in the subsequent encounters. However, the single gloss group showed an increasing trend of skipping the targeted word across the three encounters, and a decreasing trend in access and transfer the targeted word meaning, while the multiple-choice gloss group showed the opposite trend, increasingly noticing the targeted words and rehearsing their meanings across the subsequent encounters. Finally, the multiple-choice gloss group used almost twice as many strategies as the single gloss group.

The researcher interpreted these findings in relation to the Involvement Load Hypothesis, concluding that (a) lexical intervention tasks may induce different processing behavior and involvement with a new word; and (b) search and evaluation processes may trigger form-meaning connections that result not only in more robust entries in the mental lexicon but also in the encoding of more word aspects. However, these findings should be taken with caution due to the small sample of participants. In addition, there was no control group, and time on task and outside exposure between immediate and delayed posttests were not controlled.

Although some of the studies reviewed above refer to the Involvement Load Hypothesis in their discussions (e.g., Rott, Williams & Cameron, 2002; Rott & Williams, 2003; Rott, 2005), to date only four studies have attempted to directly test the predictions made by this hypothesis (Hulstijn & Laufer, 2001; Kim, 2008; Keating, 2008; Martinez-Fernández, 2008a).
Hulstijn and Laufer (2001) conducted two parallel experiments in two countries to empirically investigate whether retention of unfamiliar words was conditional upon the degree of involvement in processing these words. In this study, three intact classes of advanced university learners of English in Netherland (N=87) and three intact classes with similar participants in Israel (N=99) were randomly assigned to one of three conditions varying in the involvement load induced by the task completed: (a) in the gloss condition [+N, -S, -E], participants read a text with L1 marginal glosses for 10 targeted words, and answered ten multiple-choice comprehension questions; (b) in the fill-in condition [+N, -S, +E], participants read the same text and answered the same questions, but the targeted words were deleted from the text, leaving 10 blanks; in order to fill in the blanks they had to choose a word from a list that contained 15 words with their L1 translations and L2 explanations; (c) in the writing condition [+N, -S, ++E], participants were asked to write a composition using the targeted words, for which grammatical category, L2 explanation, example, and L1 translation were provided. Vocabulary retention was measured by unannounced immediate and delayed production posttests in which participants provided either an L1 translation or an L2 explanation for the targeted words.

The results of a 3 x 2 ANOVA indicated that the writing condition (i.e., the condition with the highest involvement load) yielded significantly higher retention than the fill-in and gloss conditions in both experiments, while the fill-in group produced significantly higher retention than the gloss condition in one experiment, but not in the other. The researchers concluded that the hypothesis was fully and partially supported in
the experiments conducted in Israel and Netherlands, respectively. However, these results can be questioned in light of the following methodological limitations: (a) experimental tasks differed not only in the degree of evaluation (-E/+E/++E), but also in input versus output orientation, and amount and quality of information provided with the targeted words in each task; (b) there was no control group; (c) process measures, such as think-aloud protocols, were not employed to ensure that tasks induced the involvement load predicted; (d) no pretest was used; instead, the likelihood of target-word familiarity was assessed in a pilot study, and prior knowledge was controlled via a post-exposure questionnaire; (e) time on task was not measured or controlled; (f) targeted items included expressions and words of different classes, so that word type was not held constant; (g) there was no randomization of participants; and (h) retention was measured by a production task only. Furthermore, participants’ level of proficiency was taken as the same in both countries, but is not adequately described; additional information about their proficiency might indicate whether this factor played a role in the different results found in each experiment (i.e., significant versus no significant difference between [+N, -S, -E] and [+N, -S, +E] conditions).

In order to expand Hulstijn and Laufer’s (2001) findings, Kim (2008) replicated their study in a different setting (English as a second language) across two proficiency levels, and controlled for time on task. Results showed that the effect of tasks was the same for both proficiency levels, with the composition group outperforming the fill-in and gloss conditions on immediate and delayed posttests, and the fill-in group outperforming the gloss group on the delayed posttest but not on the immediate posttest.
Similarly, Keating (2008) compared the effectiveness of gloss, fill-in and sentence writing conditions. Findings of this study confirmed the Involvement Load Hypothesis, but time on task played a role in these results. Although these studies controlled for proficiency level and time on task, they replicated all other methodological limitations found in Hulstijn and Laufer (2001).

These issues were addressed by Martínez-Fernández (2008a). In this study, ninety English-speaking participants enrolled in college-level second-year Spanish language courses completed a pretest, and were randomly assigned to one of four conditions that differed in whether or not the input-oriented task included need, search, and evaluation components: multiple-choice gloss condition, which involved need, search and evaluation of the meaning of the targeted words (+N, +S, +E); a fill-in condition, which involved need and evaluation (+N, -S, +E); a single gloss condition, which included only the need component (+N, -S, -E), and a control condition. The text was an adaptation of the text used by Rott (2005). In all conditions the targeted words, four concrete and four abstract unfamiliar nouns, occurred four times in the text but were glossed or deleted only in its first occurrence. The multiple-choice gloss task provided three possible translations and a ‘don’t know’ option. All participants thought aloud while reading, and completed a written retelling task immediately after. Immediate and one-week-delayed posttests measured production and recognition of the targeted words, and the ability to use them in a sentence. The think-aloud protocols revealed that the experimental conditions led to the involvement load predicted. In addition, incidental learning was controlled by a post-
debriefing questionnaire that asked learners whether they expected to be tested on the
glossed words or not.

Contrary to what the Involvement Load Hypothesis predicts, results from a final
sample of 45 participants showed that the fill-in task led to the highest vocabulary
development in all posttests (significantly outperforming both the multiple-choice gloss
and control groups), and the highest awareness. There was no significant difference
between the fill-in and single gloss conditions in vocabulary development. However,
frequency of the targeted words might have played a role in these results. Therefore, the
question of whether fill-in conditions are more conducive to vocabulary development
than single gloss and control conditions where targeted words appear only once remains
unanswered. Finally, findings on text comprehension reflected to some extent the trends
found in vocabulary development. There was no significant difference between groups in
amount of global ideas (i.e., ideas not expressed by the targeted items) recalled but the
single-gloss group significantly outperformed multiple-choice gloss and control groups
on amount of local ideas (i.e., ideas expressed by the targeted items). Interestingly, even
though the fill-in task required focusing on word form, word meaning and text content
simultaneously to a greater extent than the single gloss condition, there was no significant
difference between these groups in either local or global text comprehension.

None of the previous studies employing fill-in tasks controlled for the relationship
between the targeted word that was necessary to fill in a particular blank and the alternate
words that could be chosen to fill in that blank. In addition, Martínez-Fernández (2008a)
did not control for potential reactivity of think-aloud protocols on vocabulary
development. These aspects were addressed in Martínez-Fernández (2008b, 2009b), who examined the effect of two fill-in tasks holding the same degree of involvement load, but varying in the degree of reasoning required to choose the correct fill-in word out of two options given. In this study, 147 English-speaking adults enrolled in college-level second-year Spanish language courses were randomly assigned to one of five groups, four groups varying in the degree of reasoning required and the presence or absence of thinking aloud while reading, and one control group, in which participants were asked to read the text without glosses or blanks, and think aloud. Experimental items consisted of 12 targeted and 12 alternate words, half concrete and half abstract. The analysis of data from a final sample of 71 participants showed that (a) performance on meaning recognition, meaning production and text comprehension as measured by an open-ended questionnaire was not significantly different between the fill-in conditions, (a) high reasoning conditions significantly outperformed the control group in both meaning production and recognition immediate posttests, while the low reasoning conditions did not, and (c) task demands led to different type of processing. Therefore, future studies employing fill-in tasks embedded in reading comprehension tasks should control for the relationship between targeted and alternate words, since this variable may play a role in vocabulary learning. With respect to text comprehension, however, both fill-in conditions significantly outperform the control group on comprehension of local ideas, and no significant difference between groups was found on comprehension of global ideas.

To summarize, previous research seems to confirm that traditional glosses have a positive effect on vocabulary learning, although the vocabulary gain is generally low
(Hulstijn, 1992; Hulstijn et al., 1996; Watanabe, 1997). However, whether other procedures and lexical interventions may facilitate vocabulary learning more than traditional glosses is still a controversial issue.

Overall, two types of lexical procedures involving glossing have recently gained interest in the literature: glosses embedded in multiple-choice tasks, and glosses embedded in fill-in-the-blank tasks. With respect to the former, findings are not conclusive. Some studies comparing effects of multiple-choice gloss and control conditions, for example, did not find a positive effect for multiple-choice glosses (e.g., Hulstijn, 1992; Martínez-Fernández, 2008a) while others did (e.g., Watanabe, 1997; Rott, Williams & Cameron, 2002); similarly, when compared to single gloss conditions, some studies did not find evidence for the superiority of multiple-choice glosses (e.g., Hulstijn, 1992; Watanabe, 1997; Martinez-Fernández, 2008a), while others indicate higher performance of this condition only on delayed posttests (e.g., Rott, 2005). A number of methodological differences, such as the frequency of the targeted items and the number of options given in multiple-choice glosses, may explain these mixed results.

Studies investigating the effectiveness of fill-in tasks have found that this condition has a beneficial effect when compared to control conditions (Martínez-Fernández, 2008a) although this effect may depend on the degree of reasoning required by the fill-in task (Martínez-Fernández, 2008b, 2009b). Some studies comparing glosses embedded in fill-in tasks and traditional glosses have found a positive effect for fill-in tasks (Hulstijn & Laufer, 2001; Keating, 2008) while others support a beneficial effect only on delayed posttests (Kim, 2008), and others do not find evidence to support a
higher effect of fill-in tasks over traditional glossing (Hulstijn & Laufer, 2001; Martínez-Fernández, 2008a).

Finally, many research studies did not measure text comprehension or did not report these findings (e.g., Hulstijn, 1992; Hulstijn et al., 1996; Hulstijn & Laufer, 2001). Some researchers have argued that simultaneous attention to both form and meaning may have a detrimental effect on L2 learning, text comprehension or both (cf. Han et al, 2008). For this reason, it is important to test whether different lexical procedures may negatively impact text comprehension, have a beneficial effect or have no effect.

Among the studies that did measure text comprehension, some found that multiple-choice gloss and fill-in conditions did not impact text comprehension, as they were not significantly different from control conditions (Watanabe, 1997; Martínez-Fernández, 2008a; Rott, Williams & Cameron, 2002). In contrast, the single gloss condition significantly outperforms the control group (Watanabe, 1997; Martínez-Fernández, 2008a), and in one study significantly outperforms the multiple-choice gloss condition too (Martínez-Fernández, 2008a). However, other studies find a positive effect for multiple-choice gloss conditions, suggesting that this task leads to a higher amount of supporting ideas recalled than the single gloss condition (Rott, 2005). Additionally, there is some evidence that fill-in tasks may positively impact text comprehension when compared to control conditions (Martínez-Fernández, 2009b).

In sum, these findings suggest that the effectiveness of lexical procedures different from traditional glossing in L2 learning and text comprehension is still a controversial issue that needs to be further investigated.
Conclusion

Studies on input enhancement have investigated different pedagogical techniques that may drive learners’ attention to the linguistic features embedded in the input. The previous survey of research has presented findings of empirical studies on the following specific techniques: (1) textual enhancement, which refers to typographical or visual enhancement of the input, such as color-coding, boldfacing, highlighting, or manipulations of the font and the character size, in written texts, (2) glossing, a technique that consists of providing a definition, synonym or translation for a particular obscure word at the bottom or margin of a written text, and that usually involves typographical enhancement of the targeted item and/or the gloss, and (3) types of glossing that include an additional task, such as glosses embedded in multiple-choice and fill-in tasks.

These techniques differ in the following aspects: (a) textual enhancement has been employed mainly with grammatical items, while glossing has been typically used with vocabulary items, (b) textual enhancement is more implicit than traditional glossing because glossing provides information about the targeted items (i.e., their meaning), and (c) simple textual enhancement and traditional glossing are more unobtrusive and incidental than other procedures where learners’ attention to the targeted items is not only driven by typographical enhancement (i.e., bolding of the targeted item and/or the gloss) but also by some additional task.

The review of the previous literature on textual enhancement and different types of glossing points to minimally three areas that future studies need to address in relation to type of exposure: cognitive processes, effectiveness on L2 learning, and type of
linguistic item. First, very few studies in the textual enhancement and glossing research strands have attempted to measure attention and noticing, despite the fact that the effectiveness of these techniques depends on learners’ attention to the input. Moreover, researchers suggest that different types of glossing may lead to different types of processing (e.g., Hulstijn & Laufer, 2001). However, there is no agreement on what types of exposure to the input lead to the processing that is considered crucial for language learning. Therefore, future studies still need to investigate the effect of different types of input enhancement on cognitive processes. In addition, because studies in the cognitive psychology field indicate that levels of processing induced by different types of exposure or tasks may impact recognition memory, the literature on input enhancement may be expanded by evaluating the effects of different types of input enhancement on this dependent measure.

Second, studies on textual enhancement and traditional glossing have generally found low gains in L2 learning. For this reason, researchers have started to investigate other procedures that might lead to higher learning. However, findings on the effectiveness of different types of glossing for L2 learning and text comprehension are not conclusive. While there is evidence to support that traditional glossing has a beneficial effect on text comprehension and vocabulary learning, it is not clear whether vocabulary learning may benefit to a greater extent from types of glossing that are more language-focused. Moreover, research studies have yielded mixed results regarding the impact of these types of glossing on text comprehension.
Finally, the role of linguistic item has received little attention in glossing studies. Since different types of linguistic items may be processed differently, future research should examine the effects of type of exposure for different lexical and grammatical items, and control for type of item. The next section presents a brief review of several research studies that have addressed this variable, mainly within the textual enhancement strand.
Type of Linguistic Item in Studies on Input Enhancement

A number of factors, such as saliency, frequency, communicative value, complexity of the form, complexity of the meaning, and transparency of the form-meaning relationship, play an important role in the complexity of grammatical items (cf. DeKeyser, 2005). Consequently, specific learning problems may arise with some structures and not with others, and different learning conditions may have a differential impact on different structures. To address this issue, several studies in the textual enhancement strand have investigated the role of type of linguistic item. Shook (1994) investigated the effect of two grammatical forms differing in their degree of ‘meaningfulness’ under three conditions of exposure: textual enhancement, textual enhancement plus instruction to focus on meaning, and no enhancement. The grammatical forms were the Spanish present subjunctive (i.e., a meaningful form that carries meanings of tense and aspect) and the relative pronouns (i.e., a non meaningful form). Learning was measured by production and recognition posttests of the form and function of these items. The findings indicate that learner’s intake was significantly higher for meaningful forms than for non-meaningful forms, regardless of the condition. Additionally, a main effect was found for the condition involving textual enhancement plus instruction to focus on meaning. The researcher concluded that learners might notice a meaningful form on their own, while more explicit conditions would be necessary to impact learners’ noticing of a non-meaningful form. However, noticing was not appropriately measured in this study.
Leow et al. (2003) examined the effect of type of linguistic item on comprehension, and both noticing and immediate recognition in enhanced and unenhanced conditions. The targeted items were the Spanish present perfect and present subjunctive, and differed in terms of their saliency. Findings show that type of linguistic item did not have an effect on comprehension or intake, but more salient forms were significantly more noticed than less salient forms (as measured by an online procedure), regardless of whether they were enhanced or not. However, as the researchers indicate, some of the targeted items in this study might be considered cognates, and thus further research would need to control for this factor, since familiarity and prior knowledge may impact noticing.

While the differential effect of grammatical items has attracted researchers’ attention in the last years, there is a paucity of studies investigating the role of type of item in incidental vocabulary development. Within the glossing strand, some studies focusing on the effects of glossing on vocabulary development did not report the type of items selected (e.g., Hulstijn, 1992), some employed expressions and words of different class, so that word type was not held constant (e.g., Watanabe, 1997; Hustijn and Laufer, 2001), a few targeted only one word type (e.g., Rott, Williams & Cameron; 2002; Rott, 2005), and very few attempted to control for word type, specifically word concreteness. Martínez-Fernández (2008a) isolated the effects of concrete and abstract words under four conditions of exposure, including single gloss, multiple-choice gloss, fill-in, and control conditions. The results indicated that all groups performed significantly higher on concrete than abstract nouns, and that retention was significantly lower for abstract than
for concrete nouns on all recognition and production measures. Although these results should be taken with caution due to the low number of items employed, the study suggests that different types of lexical items might have a differential impact on L2 learning.

Generally, glossing studies have not examined effects on grammatical items, with very few exceptions (e.g., Nagata, 1999; Guidi, 2009). In contrast to textual enhancement, which only provides enhancement of the forms in a text, glosses provide the meaning of forms as well as enhancement of the forms within the text. Thus, glossing might strengthen the form-meaning relationship of grammatical items, and have a greater impact than textual enhancement alone on learners’ noticing and intake. Nagata (1999) investigated glossing for both lexical and grammatical items but did not isolate type of item as a variable. This task was undertaken by Guidi (2009), who investigated the effect of glossing across three different types of item: lexical items and two grammatical items differing in their degree of inherent difficulty (Spanish present perfect and impersonal marker ‘se’). Findings of this study revealed that (a) there was no difference between noticing of lexical items and the easy grammatical form (i.e., present perfect), while noticing was significantly higher for lexical items than for the difficult grammatical form (i.e., impersonal marker), and (b) glossing had an effect on immediate recognition of the easy grammatical form and on delayed recognition of vocabulary items. The researcher concluded that a frequency effect (multiple exposure to one grammatical form versus single exposure to multiple lexical items) might have interacted with noticing and recognition of the targeted items. Overall, this study did not support a positive effect for
glossing on L2 learning. Therefore, further research needs to establish whether different types of glossing may be effective for lexical and grammatical development, and account for the effects of type of linguistic item.

It is generally accepted that lexical items are easier to process than grammatical items because lexical items are more salient, less abstract (i.e., lexical items do not refer to functional concepts such as tense, aspect, case, etc.) and involve less complex relationships between form and meaning (DeKeyser, 2005). Within the input processing framework, VanPatten (1994, 1996, 2004) has argued that learners have a priority for processing lexical items over grammatical items when they are exposed to the input. To date, however, there is not enough empirical evidence to support this claim (cf. Leow et al., 2008 for a discussion on this issue). From a different perspective, Schmidt (1990 and elsewhere) has claimed that noticing is necessary for learning both L2 vocabulary and grammar. Very few studies, however, have attempted to investigate the effects of type of linguistic item on noticing. In sum, despite the debate of whether grammar and vocabulary stimulate different or similar cognitive processes in both first and second language, this issue has not received enough attention in empirical research, specifically in the glossing strand. Whether learners notice and remember lexical and grammatical items to the same extent, and whether grammatical and lexical development may benefit differently from different types of glossing, are issues that need to be further investigated.
Rationale and Research Questions

Rationale for the Present Study

Research in cognitive psychology suggests a relationship between awareness and experiences of recognition memory (Tulving, 1995). Studies have shown that deep processing (operationalized by different types of tasks) lead to significantly higher remembering, an experience of recognition memory that involves recollection of contextual details and that is related to episodic memory. Moreover, research studies indicate that different types of linguistic items (e.g., words versus non-words) may also have a differential impact on memory experiences, and some researchers point out that remembering plays a role in learning. Attentional models in SLA suggest that levels of awareness lead to differences in L2 learning. However, the effect of awareness on recognition memory experiences has not been investigated in the context of L2 learning. Thus, building upon cognitive psychology and SLA research, the present study seeks to investigate the effect of levels of awareness on experiences of recognition memory and L2 development.

Additionally, the SLA literature on glossing suggests that different types of exposure to the input may affect learners’ processing, learning, and reading comprehension. However, empirical studies have generally failed to measure cognitive processes, and a number of methodological differences across studies have led to conflicting findings. Moreover, although SLA research indicates that different parts of the language are not processed alike, the effect of type of linguistic item has been generally ignored in glossing studies. Thus, within a cognitive, attentional framework, the
present study aims to contribute to previous research on glossing by investigating the effect of type of glossing (i.e., traditional L1 glosses versus L1 glosses embedded in fill-in tasks) and type of linguistic item (i.e., lexical versus grammatical) on memory experiences, levels of awareness, L2 learning, and text comprehension.

Research Questions
The present study seeks to answer the following research questions:

RQ#(1) Is there a relationship between reported experiences of recognition memory and:
   (a) learners’ recognition of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?
   (b) learners’ written production of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?
If so, will the relationship be maintained over a period of one week?

RQ#(2) Is there a relationship between the level of awareness reported during a reading comprehension task and the experience of recognition memory reported by learners for (i) unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

RQ#(3) Is there a relationship between the level of awareness reported during a reading comprehension task and:
   (a) learners’ recognition of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?
(b) learners’ written production of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

If so, is this correlation maintained over a period of one week?

RQ#(4) Does type of glossing in a reading comprehension task have a differential effect on experiences of recognition memory reported by learners for (i) unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

RQ#(5) Does type of glossing in a reading comprehension task have a differential effect on levels of awareness reported by learners for (i) unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

RQ#(6) Does type of glossing in a reading comprehension task have a differential effect on:

(a) learners’ recognition of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

(b) learners’ written production of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

If so, will the effect be maintained over a period of one week?

RQ#(7) Does type of glossing in a reading comprehension task have a differential effect on learners’ text comprehension?
RQ#(8) Does type of linguistic item (i.e., lexical versus grammatical) in a reading comprehension task have a differential effect on experiences of recognition memory reported by learners?

RQ#(9) Does type of linguistic item (i.e., lexical versus grammatical) in a reading comprehension task have a differential effect on levels of awareness reported by learners?

RQ#(10) Does type of linguistic item (i.e., lexical versus grammatical) in a reading comprehension task have a differential effect on (a) learners’ recognition, and (b) learners’ written production? If so, will the effect be maintained over a period of one week?
Overview of the research design

The present study has a pretest-posttest design. The research design included five different groups varying in type of exposure and silent versus non-silent conditions for reading. Specifically, there were four experimental groups, [- incidental, + think-aloud] (N = 16), [- incidental, - think-aloud] (N = 15), [+ incidental, + think-aloud] (N = 14), [+ incidental, - think-aloud] (N = 14), and one control group [++ incidental, + think-aloud] (N = 14). The minus incidental groups were exposed to a text with glosses embedded in a fill-in-the-blank task, while the plus incidental groups were exposed to a text with traditional marginal glosses, and the most incidental group (i.e., the control group) was exposed to a text with no glosses or external manipulations. Reactivity of thinking aloud was controlled only in the experimental conditions to maximize the chances to have a representative number of participants in each group.

Table 5 shows the independent and dependent variables investigated to answer each research question, as well as the variables between which a relationship was investigated.
Table 5. Variables by Research Question

<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Independent Variable</th>
<th>Dependent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ#1*</td>
<td>Experiences of Memory</td>
<td>Recognition, Production</td>
</tr>
<tr>
<td>RQ#2*</td>
<td>Levels of Awareness</td>
<td>Experiences of Memory</td>
</tr>
<tr>
<td>RQ#3*</td>
<td>Levels of Awareness</td>
<td>Recognition, Production</td>
</tr>
<tr>
<td>RQ#4</td>
<td>Type of Glossing</td>
<td>Experiences of Memory</td>
</tr>
<tr>
<td>RQ#5</td>
<td>Type of Glossing</td>
<td>Levels of Awareness</td>
</tr>
<tr>
<td>RQ#6</td>
<td>Type of Glossing</td>
<td>Recognition, Production</td>
</tr>
<tr>
<td>RQ#7</td>
<td>Type of Glossing</td>
<td>Text Comprehension</td>
</tr>
<tr>
<td>RQ#8</td>
<td>Type of Linguistic Item</td>
<td>Experiences of Memory</td>
</tr>
<tr>
<td>RQ#9</td>
<td>Type of Linguistic Item</td>
<td>Levels of Awareness</td>
</tr>
<tr>
<td>RQ#10</td>
<td>Type of Linguistic Item</td>
<td>Recognition, Production</td>
</tr>
</tbody>
</table>

* Variables are not considered as independent or dependent variables in this case, since correlation analyses were conducted to answer the research question.

Participants

Participants were 73 English-speaking adults (41 females and 32 males) enrolled in second-year Spanish language courses. From an original pool of 135 students, 62 participants were eliminated for any of the following reasons: (1) having demonstrated prior knowledge of one or more target lexical items or prior knowledge of the grammatical structure (i.e., answering correctly more than 30% of questions on the target grammatical structure in the grammar pretests\(^\text{10}\)), (2) not possessing any prior knowledge of the past tense in Spanish (i.e., answering correctly less than 40% of questions on the past tense in the grammar pretests\(^\text{11}\)), (3) failing to follow the instructions during the

\(^{10}\text{In the present study, an attempt was made to have completely naïve learners. The cut-off was set at 30% in order to include a few participants who answered correctly 30% of questions on the grammar recognition pretest, and have a representive number of participants per cell.}\)

\(^{11}\text{Lack of prior knowledge of a simple grammatical structure, such as the past tense, might make it more difficult for learners to learn the complex grammatical structure targeted in this study. Although most}\)
treatment (e.g., when participants completed the tasks in different order, or when they tried to guess answers instead of evaluating options in the condition that required them to evaluate options), (4) failing to attend all sessions of the experiment, (5) having produced inaudible think-aloud protocols, or (6) having received formal instruction in any other Romance language.

In addition, according to the Spanish courses syllabi used in the university where data collection took place, participants had received previous formal instruction in the present subjunctive forms but not in the specific function of the present subjunctive targeted in this study. Finally, an attempt was made to control for outside exposure to target items during the experimental period by a post-debriefing questionnaire.

**Materials**

**Targeted items**

The experimental items were ten lexical items and ten grammatical items that illustrated the same grammatical structure. Two pilot tests were conducted to test learners’ prior knowledge.\(^{12}\)

The target lexical items included (a) five common, countable, and concrete nouns that referred to some relevant aspect of the main characters in the story, such as their job, place, age, or appearance (i.e., soto, ‘grove’; arce, ‘maple tree’; pordiosero, ‘beggar’; zagal, ‘young man’; orfebre, ‘goldsmith’); and (b) five regular verbs in the infinitive form, that is, a non-inflectional form, that referred to main actions taking place in the

---

12 Based on the results of the first pilot study, major changes were applied to the research design. Consequently, a second pilot test that included several new items and features was conducted.
story (i.e., descollar, ‘to be outstanding’; podar, ‘to prune’; desdénar, ‘to scorn’; embaucar, ‘to trick’; tasar, ‘to value’). Ten additional lexical items, five nouns (i.e., muelle, ‘landing pier’; tapia, ‘fence’; pijo, ‘posh person’; carcamal, ‘old person’; zahorí, ‘fortune-teller’) and five verbs (i.e., solazarse, ‘to have fun’; rociar, ‘to sprinkle water’; elogiar, ‘to praise’; hallar, ‘to find’; bruñir, ‘to polish’), were selected to provide alternate options in the text containing a fill-in-the-blank task. These items are referred to as ‘alternate items’ in contrast with ‘target items’.

The following criteria were utilized to select the lexical items: low frequency and non-cognateness. Davies (2006) dictionary of frequency of Spanish has been used to select items that did not belong to the 5,000 most frequent Spanish words. Non-cognateness was assessed through a questionnaire administered to 12 native speakers of English enrolled in a Spanish language course for absolute beginners. This questionnaire followed the format of a questionnaire used by Guidi (2009): native speakers were asked to rate on a scale from 0 (‘I can’t think of any word in English similar to this word’) to 5 (this word reminds me of the English…) the degree of cognateness of the form. Finally, all items had a clear equivalent in English so that word forms were connected to meanings in a one-to-one fashion. The English equivalents provided in the glosses were common American English words that fitted the context, and which were tested with 2 American students.

The grammatical structure targeted in this study was the use of the subjunctive in adverbial clauses introduced by the conjunction cuando (‘when’). The present subjunctive in this type of clauses is used to express future actions as illustrated by the
example in (i), while the imperfect subjunctive is used to express future actions in relation to a reference point in the past, as shown in (ii).

(i)   Iré a la playa cuando *termine* mi trabajo
      ‘I will go to beach when I *finish* my paper’

(ii)  (Le dije que) Iría a la playa cuando *terminara* mi trabajo
      ‘(I told him that) I would go to the beach when I *would finish/finished* my paper’

While participants had been formally exposed to other uses of the subjunctive in the Spanish language class, this use was expected to be unfamiliar to them. The specific form selected for the first pilot test was the imperfect subjunctive.\(^\text{13}\) Ten English-speaking adults enrolled in second-year Spanish language courses participated in the first pilot study. The results of the pilot study suggest that the structure selected might be too difficult for the participants for several reasons: (a) the use of the present subjunctive to express future actions in relation to the present time was unfamiliar to them; thus, the use of the imperfect subjunctive to express the future in relation to a reference point in the past seemed to increase the degree of difficulty; (b) even if they could understand the use of the imperfect subjunctive (one participant did), the think-aloud protocols revealed that this use was very hard to verbalize, since the notion of time was relative and not absolute; (c) while participants had been formally exposed to the present subjunctive form, most of them had not been exposed to imperfect subjunctive form; therefore, not only the use but

\(^{13}\) The target form initially selected for the pilot study was the present subjunctive. However, this form was being used as the target form in another experimental study conducted at the same time with the same participants. To avoid any potential contamination to the other study, materials of this study were modified to include the imperfect subjunctive as the target form.
also the form was unfamiliar to most of them. Since one of the goals of this study was to address levels of reported awareness of the target structure, the selection of the present subjunctive was considered more appropriate. This structure was pilot-tested with five American students studying abroad in Spain. The results of this pilot study suggested that it was easier for participants to verbalize the use of the present subjunctive although four out five participants had some prior knowledge of the target structure. After completing the grammar recognition test, participants were asked to explain why they chose one option or the other when answering each question. Interestingly, two participants were applying a wrong rule: they would select the future tense when there was a specific future time expression (e.g. tomorrow, next year...), and the present subjunctive when the future time was “general”, that is, when there was no specific future time expression. Based on this feedback, the testing items in the final study were modified to include specific future time expressions in all cases in order to prevent participants from selecting the right answer for the wrong reason.

The complexity of the Spanish grammatical structure selected (i.e., present subjunctive to express the future) for English-speaking L2 learners comes from the following characteristics: (a) the same meaning is encoded in different ways in the L1 and the L2 (i.e., there is no subjunctive form in the L1); (b) the same form may encode multiple meanings in the L2 (i.e. the present subjunctive is used to express different notions in Spanish); and (c) the same meaning (i.e., expression of future actions) may be encoded in different forms in the L2 (i.e., present subjunctive versus future tense). Typically, L2 learners generalize the use of the future tense to the adverbial clauses
introduced by the conjunction *cuando* (‘when’). Participants in this study had received formal instruction in the present subjunctive to express other notions (e.g., negative commands). However, the meaning and context in which the target structure is used was expected to be unfamiliar to them.

Finally, the grammatical items selected were familiar verbs with high frequency. Verb familiarity was assessed through a translation test that ten participants completed during the first pilot study. Only verbs that were correctly translated by 100% of participants were selected for the final materials.

**Experimental Texts**

The experimental texts in the first pilot study were adaptations of two fairy tales written by Jorge Bucay. These texts were 471 and 569 word-long, respectively, and each of them contained half of the target grammatical items and half of the target lexical items. The passages were selected because they provided a clearly developed story line that was easy to follow, and the stories were culturally neutral. However, the results of the post-debriefing questionnaire administered in the pilot test indicated that participants perceived the shorter text as very easy, and the longer text as neither easy nor difficult.

Based on the results of the first pilot study, two different texts were further pilot-tested and selected for the final study. The experimental texts were adaptations of two fairy tales written by Jorge Bucay, *El leñador esforzado* and *El verdadero valor del anillo*. Both texts were 458-word long, they provided a story line that was easy to follow, and both stories were culturally neutral. The results of the second pilot study indicated
that participants perceived a similar degree of difficulty in both texts. Finally, one text contained only target grammatical items and the other text contained only target lexical items instead of both types of items. This change to the previous design aimed to facilitate learners’ focus on the grammatical structure so that the chances to find different levels of awareness were maximized.

**Operationalization of Type of Glossing**

Three glossing conditions were operationalized in the present study: (a) L1 translation glosses (i.e., traditional glossing), (b) L1 translation glosses embedded in a fill-in-the-blank task, and (c) no gloss (i.e., control). In both experimental conditions, half of the glosses were presented at the right margin and the other half at the left margin. In the traditional glossing condition, the targeted items appeared bolded both within the text and in the glosses, as is usually done in pedagogical textbooks. Therefore, in the present study glossing involved both textual enhancement and provision of glosses. In the fill-in condition, both the targeted item and an alternate item appeared bolded within the text and in the glosses, and participants were asked to underline the item that fit better the context. Both options given in the fill-in task initially made sense, but only one was correct taking into account specific contextual clues. This version of the fill-in task was different from versions used in previous research since filling in the blank did not involve writing. Previous studies suggested that this task led to higher learning because it induced
evaluation of the items within the context and deeper processing. Therefore, the new version aimed to control for the potential effect of production.14

The results of first pilot study indicated some directions in which the fill-in task should be modified. First, with respect to the target lexical items, the fill-in task employed was too easy, since most participants selected the correct item for all blanks. In contrast to previous studies where the contextual clues were placed after encountering the blank (e.g., Martínez-Fernández, 2008a, 2009b), in this case the contextual clues were placed immediately before the blank. The think-aloud protocols revealed that participants did not need to engage in evaluation processes and reconsider previous answers, as has been found in previous studies. Thus, this task was modified to increase the chances to find different levels of processing.

With respect to the target grammatical items, the choices given to fill in the blanks were two different familiar verbs with (a) similar meaning (i.e., both options could work in the context), and (b) different tense form (i.e., only one was correct in the context). The think-aloud protocols showed that participants often felt that one verb fit better the context than the other, which led them to choose one verbal form because of the verb meaning without paying attention to the verbal form. Like the choices for lexical items, the alternatives for grammatical items were also modified in order to increase the saliency of the targeted structure. Instead of presenting two different verbs, the choices in fill-in task used in the final study presented two different forms of the same verb: the

---

14 Although the task used in the present study required participants to choose one of two options instead of filling in a blank, it will be referred to as a fill-in task in order to facilitate comparisons across studies that address the effect of evaluation on vocabulary learning (e.g., Hulstijn & Laufer, 2001).
present subjunctive and a past tense form. To select the correct option, participants would have to evaluate the context, which would make clear the need for a non-past tense form. Besides the ten target grammatical items, five distractors were included in the final task. The alternatives given for the distractors were the same (i.e., present subjunctive and past tense) but the correct option was the past tense, and not the present subjunctive. The distractors were included to motivate further evaluation of the grammatical forms.

Both fill-in-the-blank tasks (i.e., with lexical versus grammatical items) were pilot-tested by 10 native speakers of Spanish, who were asked to fill in the blanks and indicate the contextual clue that has led them to choose one option or the other in each case.

Finally, the no gloss condition involved reading the experimental texts without any kind of external enhancement. Therefore, this condition will be referred to as the most incidental condition (++ incidental). The traditional glossing condition was considered less incidental (+ incidental) because it aimed to drive readers’ attention to the targeted items by enhancing lexical and grammatical items, and providing their meaning. The fill-in condition was considered the least incidental condition (- incidental) since it aimed to drive learners’ attention to the targeted items by enhancing them, providing their meaning, and asking them to distinguish the targeted item from the alternate items provided.
Assessment Measures

Process measures

This study included think-aloud protocols in order to (a) assess whether participants were representative of their experimental condition, that is, they fulfilled what was required of them in this condition, (b) measure attention and awareness while reading the experimental texts, and (c) measure experiences of remembering and knowing. Ten per cent of the think-aloud protocols produced while reading and ten per cent of the think-aloud protocols produced while completing the remember-know task were randomly selected to be analyzed by two researchers and obtain inter-rater reliability (Kappa coefficient was high in both cases: .95 and .94, respectively).

The analysis of concurrent verbalizations allowed the researcher to determine whether participants engaged in the reading of the texts in all conditions, whether they paid attention to the glosses in the experimental conditions, and whether they evaluated different options to fill in the blanks in the minus incidental condition. The following quotes illustrate how participants in different groups engaged in different processes:

(a) Participants in the control condition

“I have never been able to… descollar… I don’t know descollar…”

“Un día al salir del trabajo [skips target item] un compañero…”

(b) Participants in the plus glossing condition

“Un día al salir del trabajo en el soto, grove, a friend…”

“He had never been able to be outstanding”
(c) Participants in the minus glossing condition

“The merchants began to… praise or scorn… praise the ring [continues reading] Oh, maybe scorn because no one bought the ring…”

“I have never been able to have fun, be outstanding… [continues reading]. How can I get better? Yes, be outstanding…”

“I don’t think a maestro is a posh person but I don’t know… [continues reading] camisa sucia y rota… so he looks like a beggar, not like a posh person, so pordiosero”

Overall, participants used a variety of strategies while reading, ranging from reading aloud sentence by sentence in Spanish and translating immediately after, translating word by word directly, reading in silence and summarizing main ideas, and using any of the strategies above plus making personal comments about the content of the passage (cf. Leow et al., 2008 for similar findings).

In order to measure instances of noticing and awareness, a coding procedure was established after a preliminary analysis of the data collected, based on previous studies employing think-aloud protocols. For lexical items, each item was assigned a level of reported awareness. Initially, data were coded according to the following categories: no verbal report (e.g., no mention of a target item, or reading aloud a target word without making any pause); noticing of form (e.g., any mention of a target form accompanied by a pause or by a mention of not knowing the meaning of the form); noticing of meaning
(e.g., any mention of the translation or the meaning of a target item); noticing of both meaning and form (e.g., any mention of a target form accompanied by a mention of its translation or meaning); and awareness at a higher level than noticing (e.g., any comment about the targeted items that indicated an explicit focus on the items as language *per se*).

The following report illustrates the only case of awareness at a higher level than noticing that was found in the data:

“This is new vocab, *za-gal*, I had no idea, ‘young man’” (Participant in the plus incidental condition)

Because there were no other instances of awareness at a higher level than noticing, this report was coded as noticing of both meaning and form. In order to analyze the data, different types of noticing (i.e., noticing of meaning versus noticing of form versus noticing of meaning and form; and noticing of one word aspect versus noticing of two word aspects) were taken into account. However, because the results indicated no differences between types of noticing, all categories were conflated into one (i.e., noticing), following Guidi (2009). Therefore, reports on lexical items were reinterpreted into two categories: no verbal report and noticing. The following examples illustrate the category of noticing:

---

15 In this study, reported noticing of one word aspect and reported noticing of two word aspects were found to a similar extent in plus and minus incidental conditions, in contrast to findings of previous research (e.g., Martínez-Fernández, 2008a).
“Soto… grove… because landing pier doesn’t make sense” (Participant in the minus incidental condition)

“… leaving work in the grove” (Participant in the plus incidental condition)

“One day leaving work in el soto, no sé soto…” (Participant in the control group)

In order to measure attention and awareness of the grammatical structure, two different measures were used: (1) online measures, and (2) a combination of online and offline measures. For online data, each participant was assigned three scores: (a) number of instances of no verbal report (e.g., no mention of a target item, or reading aloud a target item without making any pause); (b) number of instances of noticing (e.g., any mention of a target item accompanied by a pause and/or its translation); and (c) one or zero point for reported understanding depending on whether the participant had verbalized the rule underlying the use of the target grammatical forms at least once (i.e, one point) or not (i.e., zero point). The following are examples of understanding and noticing as measured by think-aloud protocols:

(a) Understanding:

“Cuando la próxima semana… venga… ok, it is in the future, next week, so it’s not vine, that would be stupid… So venga is used to express the future… Yes, the present subjunctive” (Participant in the minus incidental condition)

“When I sleep the next day… so it’s in the future… it’s the present subjunctive, not the past” (Participant in the minus incidental condition)
(b) Noticing:

“When tomorrow I… find out… descubra the conditions” (Participant in the minus incidental condition)

“Cuando mañana por la noche… when tomorrow night… I am able to say that I cut 25 trees in one day…” (Participant in the plus incidental condition)

Additionally, participants completed a post-exposure questionnaire in which they were asked to indicate (a) whether they had noticed any grammatical form in particular while reading the passage, and specify which one, and (b) whether they had learned any grammatical rule, and specify which one. Although offline procedures to measure awareness are controversial, as has been explained in the literature review, offline data in this study were useful to confirm online data, on the one hand, and to have at least some indication of the level of awareness reached by those participants who were not in the think-aloud groups, on the other. The following are examples of understanding and noticing as measured by the post-exposure questionnaire:

(a) Understanding:

“Yes, the subjunctive is used when referring to possible future events”

( Participant in the minus incidental condition)

“Yes, that the subjunctive is used when you talk about when (cuando) something will happen in the future” (Participant in the plus incidental condition)
“Yes, if you say ‘I will do ________ tomorrow’ you still use the future tense, not the present. If you say ‘He will be happy when I come tomorrow’, he will be happy is in the future tense but I come would be in the present subjunctive even if it is a future event” (Participant in the plus incidental condition)

(b) Noticing

“Yes, I noticed the present subjunctive and future tense used often” (Participant in the minus incidental condition)

“Yes, I think they used the present subjunctive a lot” (Participant in the plus incidental condition)

Instances of understanding and noticing of the grammatical structure were found in all experimental conditions (i.e., plus and minus incidental conditions, and think-aloud and silent conditions) although cases of reported understanding were found less often overall (the data revealed only ten cases). Based on the offline and online data, participants were assigned to one of the following categories: (a) understanding (i.e., if the participant verbalized the rule underlying the target grammatical forms either during the think-aloud protocols or in the post-exposure questionnaire), (b) noticing (i.e., if the participant reported noticing of at least 60% of the target items as measured by the think-aloud protocols, or indicated to have noticed the present subjunctive while reading in the post-exposure questionnaire), and (c) low amount of verbal report (i.e., if the participant reported noticing of less than 60% of the target items as measured by the think-aloud protocols, and indicated not to have noticed any particular grammatical form while
reading in the post-exposure questionnaire). All participants who reported noticing of less than 60% of the target items during the think-aloud protocols indicated not to have noticed any grammatical form in particular in the post-exposure questionnaire. Because classifying these participants into the ‘noticing category’ would contradict results from the post-debriefing questionnaire, and classifying them into the ‘no verbal report’ category would contradict results from the think-aloud protocols, a decision was made to create a new category, that is, ‘low amount of verbal report’, that would account for results from the two measures.

Awareness of lexical items as measured by an online procedure was considered as a within-subject variable (i.e., each participant in the think-aloud conditions was assigned different scores for noticing and no verbal report). Similarly, awareness of the grammatical structure as measured by an online procedure only was considered as a within-subject variable (i.e., each participant in the think-aloud conditions was assigned different scores for understanding, noticing, and no verbal report). In contrast, awareness of the grammatical structure as measured by both online and offline procedures was considered as a between-subjects variable in the statistical analyses (i.e., participants in think-aloud and silent conditions were assigned to different groups: understanding, noticing, and low amount of verbal report). Data on awareness of the grammatical structure as measured by (1) an online procedure, and (2) a combination of online and offline procedures were submitted to separate statistical analyses in order to answer the research questions addressing the role of levels of awareness of grammatical items. Since an online procedure was considered to be a more robust measure of awareness, the
Results section addressed primarily results of online data; results of the analyses on online plus offline data were reported only when they differed from the analysis on online data and added a new insight in the findings.

As mentioned above, think-aloud protocols were also used to measure experiences of remembering and knowing. Data on memory experiences were analyzed to increase the internal validity of the task used in cognitive psychology, which is described below within this Assessment Measures section.

Finally, the issue of reactivity was addressed by controlling the effect of thinking aloud in the experimental groups. In addition, the post-debriefing questionnaire of the pilot study included two questions asking participants whether it had been difficult for them to think aloud while reading and while completing the recognition memory task. With respect to think aloud while reading, 59% of participants (N = 17) answered “no,” 12% answered “a little” or “sometimes,” and 29% answered “yes.” With respect to the verbalizations produced while completing the post-exposure memory task, 83% answered “no,” 6% answered “sometimes,” and 12% answered “yes.” These results suggested that overall most participants found it easy to think aloud, and found it easier to think aloud while completing the memory task than while reading. This difference across tasks might be due to the fact that the reading task was more cognitively demanding, or to a practice effect, as some participants indicated that it was hard to think aloud at the beginning but not later. The following quotes illustrate answers from participants who indicated that thinking aloud while reading was not difficult:
“No. I often talk to myself when reading or doing other school work”

“No, this is usually how I work on Spanish assignments”

“At first, but when I began to have trouble working through the passages, it came almost naturally. Just an extension of my thought processes”

“Not really. I thought it would be but I actually didn't mind doing that at all and it seemed easy”

Interestingly, some participants who answered that it was easy seem to refer to a positive effect of thinking aloud:

“I think I might have done my thinking a little differently, tried to put more analysis in rather than simply understand what the text is saying”

“No, it made understanding the texts easier”

In contrast, the following quote reveals that the difficulty to think aloud may have a negative impact on reading comprehension:

“It was difficult because I did not know what to say and I usually read a paragraph and then think about it, or else it's harder to comprehend”

Although this comment is representative of only 12% of participants, it suggests that instructions to think aloud may affect the reading strategies adopted by the
participants, leading them to read in a bottom-up fashion and precluding them from using top-down strategies, such as recapitulating ideas and integrating new content. Based on this result from the pilot study, the instructions to think aloud were modified in the final study. The new instructions did not encourage participants “to think aloud constantly” or “not to stop verbalizing their thoughts,” but rather specified that they should think aloud only when they were actually thinking. In addition, the following quote from the pilot test data indicated the need to emphasize that any language could be used at any point while thinking aloud so that not everything needed to be translated into the L1:

“A little bit. It was sort of strange to voice out my thought process, though now I realize maybe I don't have a very vocal one. When I read in Spanish I sort of just process the words in my head in Spanish, unless then I really don't understand them, then I try to translate to English”

Finally, one participant referred to his/her inability to verbalize thoughts about a different language, and a few indicated that it was “awkward.” In sum, these qualitative data from the pilot study showed how learners approached the task of thinking aloud in different ways, and provided insightful information to improve instructions and directions. While in the pilot study participants were given an example of how to think aloud when solving a Math problem, the final study included also additional practice with reading, using a short text different from the experimental texts. Moreover, an attempt was made to control potential reactivity of thinking aloud through the post-debriefing
questionnaire so that participants who found it difficult to think aloud would be excluded from the final sample. However, the analysis of data from the final study revealed that the percentage of participants who found it difficult to think aloud was higher than in the pilot study: 44% of participants in the minus incidental condition, and 43% in the plus incidental and control conditions. Since there was no difference across groups in the percentage of participants who had trouble to think aloud, no participants were excluded from the final sample. This decision was made in order to keep a representative number of participants by cell. The reason most often mentioned by participants to explain why thinking aloud was difficult points to the fact that the laboratory where the treatment took place was crowded, and it was uncomfortable to hear other people thinking aloud. The following quotes illustrate this experience:

“Yes, it felt strange to speak out loud when other people were doing different parts of the experiment. Also, thinking out loud is difficult to verbalize. Thoughts are not always in coherent and organized sentences.” (Participant in the minus incidental condition).

“Yes, when the room was crowded it was more difficult to think out loud and concentrate with other people speaking as well.” (Participant in the plus incidental condition).
“Yes, because usually when I read, I do not think aloud unless I am really frustrated and cannot figure something out. Also, there were multiple other students in the room, so it was kind of uncomfortable to think aloud.” (Participant in the plus incidental condition).

Evidence of this complaint was also found in the think-aloud protocols. In order to understand this general complaint, it is important to take into account the unusual conditions in which the treatment took place. Most sessions scheduled for the treatment week had to be cancelled because of the historic snowstorm that hit Washington DC. Consequently, very few sessions were available to complete the treatment, and the laboratory turned out to be crowded during all of them. This circumstance might also explain why the percentage of participants who found it difficult to think aloud was much higher than in the pilot study.

Finally, 23% of participants considered it difficult to think aloud while completing the remember-know test, which confirms that thinking aloud while completing this test was perceived as an easier task than thinking aloud while reading.

**Reading Comprehension Questionnaires**

Reading comprehension was measured prior to the treatment to control for a potential effect of reading skills on post-exposure text comprehension. In the pilot study, the reading comprehension pretest was an adaptation of a multiple-choice test designed and published by Instituto Cervantes for the *Diploma de Español como Lengua*
Extranjera, a standardized exam of Spanish as a foreign language. The text used for this test was an adaptation of an article published in a Spanish newspaper, which narrated the adventures of a couple on a long trip. The reliability of the 6-item multiple-choice test turned out to be quite low (Cronbach’s alpha = .45), and therefore a different pretest was used in the final study. This new pretest was designed with similar characteristics to the tests used after exposure. A similar test was deemed more appropriate since learners’ reading skills may differ across text types (i.e., learners may be good at reading expository texts but not fairy tales), and their ability to complete a test may be different for different types of tests (i.e., they may be good at completing multiple-choice tests but not open-ended questionnaires). The reading comprehension pretest was similar to the posttests in the following aspects: the text to be read was an adaptation of a fairy tale (El círculo del 99) written by the same author as the experimental texts, and had similar length (470 word-long), and the test consisted of ten open-ended questions to test global comprehension.

Two post-exposure reading comprehension questionnaires were pilot-tested and modified to address subsequent content modifications in the experimental texts. The questionnaire designed to measure comprehension of the text containing target lexical items included (a) ten open-ended questions that addressed comprehension of global ideas (i.e., ideas not expressed by the targeted lexical items), and (b) ten open-ended questions that addressed comprehension of local ideas (i.e., ideas expressed by the targeted lexical items). The questionnaire designed to measure comprehension of the text
containing target grammatical items included ten open-ended questions to address global comprehension.

The Remember-Know Task

The remember-know task was the same as the one used in cognitive psychology studies. Prior to completing the task, participants were given written definitions and examples of remember, know and guess responses. In the task, items were presented one by one, and participants were asked to indicate whether each item was present in one of the texts or not, and give a remember, know or guess response. In addition, they were prompted to verbalize any memories they might have (e.g. if they remembered whether the word appeared in one text or the other, or on the top of the page, at the bottom or in the middle; whether the word was glossed or not; whether a gloss for a word was presented at the right or left margin; whether they remember reading, seeing, visualizing, pronouncing, underlining or thinking of that word; etc.). The task included 50 items (i.e., 10 target lexical items, 10 alternate lexical items, 10 lexical distractors, 10 target grammatical items, and 10 grammatical distractors). Distractors for lexical items were 5 nouns and 5 verbs in the infinitive form that could make sense in the context of the stories but did not appear in any of them. Distractors for grammatical items were the same ten target verbs but in a tense different from that of the targeted or alternate verbal forms.

The analysis of the think-aloud protocols produced during this task in the pilot study revealed that thinking aloud was an effective measure to control for the validity of
participants’ responses. The following comments illustrate experiences of remembering that led to remember responses:

“They used *capataz* as the definition on the side bar, meaning *foreman*, talking about Ivan”

“Yes, definitely, this is the word for *ax*, I remember thinking that because in High School in one of the vocab lessons we had vocab on camping, and I was thinking that the only time I have ever used that word ever was here in this test”

“Yes, I chose this word in the Hassan’s story…”

The comments below illustrate experiences of knowing that led to know responses:

“It was definitely in there, I don’t know what it means and I don’t know where exactly it was”

“I do not remember exactly, but I do know that it was in the story”

While the examples above reflect valid responses, participants often failed to distinguish the experience of knowing from guessing. As has been pointed out by researchers, the experience of knowing may involve different levels of confidence. Thus, when participants have a feeling of familiarity but they are not completely sure about
having encountered a word, they often interpret their experience as guessing. The comments below illustrate several experiences of knowing that led to guess responses:

“I recognize this, I think it was used but I am not one hundred per cent sure, not sure where it would have been used either cause I don’t remember what it means”

“I remember it but I have no idea what the meaning or the use of this word is… it’s a guess, I don’t think it was in the Circle story, I think in the Woodcutter one, or maybe not, I don’t know exactly”

“I don’t remember that one but it’s possible I have seen a word like copla in one of the stories but I don’t know in which one so I am guessing”

“I remember this from… I think it was the Circle 99, no, I don’t remember when or where but I recognize this word and I know that I haven’t heard it before”

Less often, participants interpreted an experience of remembering as knowing. The data revealed that the experience of remembering might also involve different degrees of strength, as has been recently suggested by some researchers. The comments below reflect that some participants could remember themselves thinking about a word, or could remember seeing a word, but could not recognize whether that word appeared in one text or the other. Thus, they could retrieve some aspects of their encounter with an item but could not retrieve others. The first comment comes from a participant who gave a know response, while the second one comes from a participant who provided a guess response:
“This… I am positive it was in there, yes, I thought about this word for a while, I am definitely sure that it was in one of the texts but I don’t know which one”
“I remember seeing this, I am pretty sure, yes, but I don’t know from which text… so I am gonna say guess”

Additionally, the think-aloud data show that in some cases participants may experience a ‘false memory’, that is, a detailed and vivid memory of something that has never occurred. A few participants from both the pilot test and the final study gave a remember response for a targeted item, and explained where exactly they had seen this word by providing details of the context. However, the item did not appear in any of the texts, or it did not appear in the text where participants remembered to have seen the word, or the context was completely different. The examples below illustrate cases of remember responses that reflect false memories for different items:

(a) *Saltar*, ‘to jump’, an item that did not appear in any of the texts:

“Yes, I remember this one, it was in the Circle 99, Hassan was jumping around everywhere!”

(b) *Hacha*, ‘ax’, an item that appeared in the Woodcutter story, and not in the Circle 99 story:

“Yes, I remember, this was when the king was firing or exiling Hassan in the Circle 99”
(c) *Viera*, ‘would see’, an item that appeared in the Circle 99 story to refer to the time
when the king would see his advisor to ask him a question:

“Yes, it was in the Circle 99, when they went to the house and were looking at
Hassan”

(d) *Comiera*, ‘would eat’, an item that appeared at the very end of the Circle 99 story:

“Yes, that's in the first one, I remember talking about that at the very
beginning of the story”

In the post-debriefing questionnaire, participants were asked whether it was
difficult for them to distinguish remember from know responses, and know from guess
responses. Overall, participants found it easy to make these distinctions. Twenty-seven
per cent of participants indicated that it was difficult to distinguish remember from know
responses, and thirty per cent indicated to have difficulty to distinguish know from guess
responses. The explanations below reflect issues that have been discussed in the
literature, and which give support to the use of think-aloud protocols to control for the
internal validity of this test:

“I think it was slightly difficult only because I felt that the two (remember and
know) were close, and I needed 100% certainty to remember while the range for
know was much larger.”

“I think there should only be a remember or a know category. In my opinion the
two are too similar. I feel as though they are very alike and depending on the
person taking the test the judgments could be different when in reality they have
the same interpretation of the words, just not of the categories to place them in.”

In sum, the think-aloud protocols may allow researchers to correct invalid
responses (i.e., responses based on misinterpretations of remembering, knowing and
guessing) and, therefore, increase the internal validity of the remember-know task. In the
present study, the retrieval of only one contextual aspect of the encounter with a targeted
item was sufficient to consider the experience as remembering, and a vague sense of
familiarity reported was sufficient to consider the experience as knowing, regardless of
confidence degree. The issue of false recognition was addressed by using the traditional
method employed in cognitive psychology: the number of remember responses given to
distractors/new items (i.e., ‘false alarms’ or false memories) were subtracted from the
number of remember responses given to target/old items (i.e., ‘hits’ or true memories),
following the same procedure for know and guess responses. Although the think-aloud
protocols were useful to detect several cases of false memories, not all reports were
detailed enough to establish the occurrence or not occurrence of a false memory. In
addition, the number of remember and know responses given to distractors suggests that
there might have been more cases of false memories than those detected in the think-
aloud protocols. Therefore, if cases of false memories as measured by think-aloud
protocols were excluded from the data, this method should have been combined with the
quantitative method used in cognitive psychology. However, the quantitative method
calculates false recognition overall, taking into account responses given to the same
number of old and new items. Thus, the combination of qualitative and quantitative
measures might have yielded misleading results in this case. For these reasons, the quantitative method was preferred in the present study.

Finally, participants in the pilot study (N = 10) were assigned to one of two groups varying in the order in which they read the texts. The results suggested that text order might affect experiences of memory, leading to higher episodic memory of both lexical and grammatical items appearing in the text read in the second place. Therefore, in the final study participants were assigned to different text order conditions within each group to control for this factor.

**Learning measures**

In this study learning is operationalized as the ability to recognize and produce the grammatical structure and the meaning of lexical items. Recognition and production tests were administered to measure intake and L2 development, respectively, immediately and one week after the treatment.

**Recognition tests (pre- and post-)**

Recognition of lexical items was measured by a multiple-choice test. In this test, participants were presented with the L2 targeted items, and were asked to recognize its meaning by choosing one out of four options: the gloss used for the target item, a gloss used for other target item (i.e., a target item of the same word class), a cognate, and an “I don’t know” option. This test included 30 items: 10 target items, 10 alternate items, and 10 distractors (i.e., the same items used as lexical distractors in the remember-know task).
Recognition of the grammatical structure was measured by a fill-in-the-blank task with multiple choices. This test consisted of 30 L2 sentences containing a blank: 10 sentences with the targeted structure and the targeted verbs (i.e., old items), 10 sentences with the targeted structure and non-targeted verbs (i.e., new items), and 10 sentences (5 with old verbs and 5 with new verbs) containing a distractor structure (i.e., temporal clauses introduced by cuando to express actions in the past). The choices given to fill in the blank in sentences containing the target structure included the correct form (i.e., present subjunctive), two incorrect forms that are typical errors (i.e., future tense and present indicative), and an “I don’t know” option. The choices given to fill in the blank in sentences containing the non-target structure were the correct form (i.e., past tense), the target form (i.e., present subjunctive), another incorrect form that expresses a more neutral time (i.e., present indicative: a non-past and non-future form), and an “I don’t know” option.

**Written production tests (pre- and post-)**

Meaning production of the lexical items was tested by an L2-to-L1 translation test including 30 items (i.e., 10 target items, 10 alternate items, and 10 distractors).

Production of the grammatical structure was tested by a fill-in-the-blank task including 30 sentences (i.e., 10 containing the target structure with old items, 10 containing the target structure with new items, and 10 containing the distractor structure). Participants were asked to write the L2 verbal form for the verbs presented in parenthesis. Distractors were the same as those used in the recognition tests.
The University Internal Review Board approved all materials and procedures.\footnote{Reference of IRB approval: 2009-577.} Participants were recruited from second-year Spanish college classes (Intermediate I and Intermediate II). A research assistant visited the classes to give participants a short description of the goals of the study and invite them to participate. Participants received a consent form and a sign-up sheet to mark the date and time when they would complete the different sessions of the study.

The assessment materials were computerized using Blackboard. This system has been used because it presents the following advantages: (a) it is linked to a database that records each answer into a ‘grade book’ that can be accessed by the researcher; (b) the system can prevent users from coming back to a previous answer or page; and (c) it does not allow participants to see all tasks of the experiment at the same time; instead, they can only see the task they have to complete, and only after completing this task they can see the next one. The experimental texts were not computerized because the system lacked the necessary features to computerize the fill-in task. However, in order to present all experimental tasks in the same mode, the texts were also be presented on the computer, and participants in the fill-in condition were asked to underline the options chosen to fill in the blanks with the computer feature for underlining.

Participants completed three sessions in the laboratory. The first session took place two weeks prior to the treatment. In this session, participants completed the reading
comprehension, grammar production, word meaning production, grammar recognition, and word meaning recognition pretests.

In the second session, participants were randomly assigned to groups. The think-aloud groups received oral instructions to think aloud while reading as well as practice with a short text. Before reading the experimental texts, all groups were informed about a subsequent text comprehension test, but not about the remember-know, word meaning production, grammar production, word meaning recognition, and grammar recognition posttests. Participants read two texts (presented in randomized order) at their own pace, and time on task was measured. Immediately after completing the first reading, they completed the text comprehension questionnaire. Then they proceeded to read the second text, and completed the text comprehension questionnaire for that text.

Once participants had completed both readings and text comprehension tests, they received oral and written instructions to make recognition memory judgments. All groups were asked to think aloud while completing the remember-know task. The remember-know task was followed by production tests, and production tests preceded recognition tests. This order was selected for two reasons: (a) to avoid a potential effect of the exposure to the targeted items provided in the production and recognition tests on the remember-know task, and (b) to avoid a potential effect of the choices provided in the recognition tests on the production tests. The remember-know task did not require participants to think of the meaning and use of targeted items so that its potential effect on subsequent posttests should be smaller than the effect of these posttests on recognition
memory. Finally, the order of word meaning and grammar production tests was randomized, as well as the order of word meaning and grammar recognition tests.

One week after the treatment, participants completed the remember-know task, followed by production and recognition tests. The same items as in immediate posttests were presented but the order was randomized. However, immediate and delayed word meaning posttests differed in whether or not they included distractors. Immediate word meaning posttests included distractors (i.e., the same distractors used in the remember-know task) in order to minimize the effect of exposure to the targeted items during immediate posttests on the delayed remember-know task completed one week after exposure. Thus, distractors did not play any role in the delayed word meaning posttests since these tests were completed after the delayed remember-know task. In other words, exposure to the targeted items in the delayed word meaning tests could not affect performance on the delayed remember-know task. For this reason, distractors were not included in the delayed word meaning posttests. In contrast, distractors in grammar production and recognition tests were used to control for other additional factors. Two types of distractors were used with different purposes: (a) non-targeted verbs were included to test whether participants were able to generalize their knowledge of the grammatical structure to different items (i.e., new items), and (b) a non-targeted structure was included to control for a potential overgeneralization of the targeted structure to

---

17 Due to the loss of almost half of the think-aloud protocols produced while completing the remember-know task during the delayed posttest, these data have not been submitted to analysis.  
18 The delayed remember-know task asked participants to indicate whether they remembered to have seen the target items in the reading passages. However, they could provide remember or know responses just because they had encountered the target items several times during the immediate posttests, and not because they remembered to have seen them in the reading passages.
other contexts (i.e., contexts requiring the past tense instead of the present subjunctive). Therefore, distractors in grammar production and recognition tests played a role both at immediate and delayed posttests, and were included in all of them.

Finally, at the end of the third session, participants completed a post-debriefing questionnaire designed to address issues such as outside exposure, participants’ perceptions, background information, etc. Figure 1 illustrates the procedure of the study, and includes the average time that each session and each test took.
Figure 1. Procedure

EXPERIMENT

WEEK 1
Session 1
(45 min. approx)

READING COMPREHENSION PRETEST
(13 min. approx.)

WRITTEN PRODUCTION PRETESTS
(in randomized order)
Vocabulary (5.52 min)
Grammar (12.04 min)

RECOGNITION PRETESTS
(in randomized order)
Vocabulary (3.51 min)
Grammar (7.25 min)

WEEK 3
Session 2
(1 h. 15 min. approx.)

TREATMENT
Reading of Text (I)
(8.11 min.)

TEXT COMPREHENSION POSTTEST FOR TEXT (I)
(6 min. approx)

REMEMBER-KNOW TASK
(preceded by oral + written instructions)
(8 min. approx.)

WEEK 4
Session 3
(30 min. approx)

TREATMENT
Reading of Text (II)
(8.11 min.)

TEXT COMPREHENSION POSTTEST FOR TEXT (II)
(6 min. approx)

POST-DEBRIEFING QUESTIONNAIRE

REMEMBER-KNOW TASK
(preceded by oral + written instructions)
(8.3 min.)

WRITTEN PRODUCTION POSTTESTS
(in randomized order)
Vocabulary (4.3 min)
Grammar (8 min)

RECOGNITION POSTTESTS
(in randomized order)
Vocabulary (3.15 min)
Grammar (5.2 min)
CHAPTER 3: RESULTS

This chapter is divided into five parts. The first part reports the results of the preliminary analyses conducted to ensure the validity and reliability of the analyses that address the research questions. The next parts report the results obtained for each of the independent variables of the study - experiences of recognition memory, levels of awareness, type of glossing, and type of linguistic item - respectively.

Part I. Preliminary Analyses

Prior to addressing the research questions of the study, several preliminary steps were taken to ensure the validity and reliability of the results. Specifically, the following methodological issues were examined: prior knowledge of the target items, reactivity of verbal reports, time on task, text effect, and reliability of dependent measures. All statistical analyses were run using the Statistical Package for the Social Sciences (SPSS), with the alpha level set at 0.05 throughout.

Pretests

To ensure that groups started out at statistically similar ability levels, several one-way ANOVAs were conducted on the pretests. Participants with prior knowledge of one or more lexical items were not included in the final sample. Thus, no statistical analysis was necessary to compare prior knowledge of vocabulary across groups. Statistical
analyses were run to compare prior knowledge of grammatical items and reading comprehension ability.

*Text Comprehension Pretest*

Finally, to ensure that groups started out at statistically similar reading comprehension levels, a one-way ANOVA was conducted on the reading comprehension pretest. The results indicated no significant differences between groups in their ability to answer open comprehension questions about a text ($F(4, 72) = 2.160, p = .08$).

Descriptive statistics (means, SD, an SE for text comprehension pretest by group) are displayed in Table 6.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-I, +TA]</td>
<td>16</td>
<td>6.44</td>
<td>1.931</td>
<td>.483</td>
</tr>
<tr>
<td>[-I, -TA]</td>
<td>15</td>
<td>4.93</td>
<td>2.344</td>
<td>.605</td>
</tr>
<tr>
<td>[+I, +TA]</td>
<td>14</td>
<td>5.86</td>
<td>2.248</td>
<td>.601</td>
</tr>
<tr>
<td>[+I, -TA]</td>
<td>14</td>
<td>7.07</td>
<td>2.129</td>
<td>.569</td>
</tr>
<tr>
<td>++I, +TA]</td>
<td>14</td>
<td>5.57</td>
<td>1.950</td>
<td>.521</td>
</tr>
</tbody>
</table>

Note. Maximum score = 10.

*Grammar Recognition Pretest*

To determine whether participants started out at similar ability levels to recognize a familiar grammatical structure (i.e., Past tense) and the target unfamiliar structure (i.e., Present Subjunctive to express the future), scores on the grammar production pretests were submitted to a one-way ANOVA. The results indicated no significant differences between groups in their ability to recognize the Past tense ($F(4, 72) = .257, p = .90$) or
the Present Subjunctive \( F(4, 72) = .137, p = .96 \) in the respective pretests. All groups were able to recognize the Past tense to a similar extent, and were unable to recognize the Present Subjunctive to express the future. Descriptive statistics (means, SD, an SE for recognition pretests of two types of items by group) are displayed in Table 7.

Table 7. Means, SD, and SE for Grammar Recognition Pretests for Two Types of Item by Group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-I, +TA]</td>
<td>16</td>
<td>6.88</td>
<td>1.821</td>
<td>.455</td>
<td>1.00</td>
<td>1.897</td>
<td>.474</td>
</tr>
<tr>
<td>[-I, -TA]</td>
<td>15</td>
<td>6.60</td>
<td>1.882</td>
<td>.486</td>
<td>1.13</td>
<td>1.685</td>
<td>.435</td>
</tr>
<tr>
<td>[+I, +TA]</td>
<td>14</td>
<td>6.93</td>
<td>1.979</td>
<td>.529</td>
<td>.93</td>
<td>1.639</td>
<td>.438</td>
</tr>
<tr>
<td>[+I, -TA]</td>
<td>14</td>
<td>7.21</td>
<td>2.082</td>
<td>.556</td>
<td>1.36</td>
<td>1.906</td>
<td>.509</td>
</tr>
<tr>
<td>[++I, +TA]</td>
<td>14</td>
<td>7.21</td>
<td>1.968</td>
<td>.526</td>
<td>1.07</td>
<td>.997</td>
<td>.267</td>
</tr>
</tbody>
</table>

Note. Maximum score for Past Tense = 10. Maximum score for Present Subjunctive = 20.

Grammar Production Pretest

A one-way ANOVA was conducted on the grammar recognition pretests to determine whether participants started out at similar ability levels to recognize a familiar structure (i.e., Past tense) and the target unfamiliar structure (i.e., the Present Subjunctive to express the future). The results of this analysis indicated no significant differences between groups in their ability to produce the Past tense \( F(4, 72) = 1.376, p = .25 \) or the Present Subjunctive \( F(4, 72) = 1.154, p = .33 \) in the respective pretests. Therefore, all groups had a similar ability to produce the Past tense, and a lack of ability to produce the Present Subjunctive to express the future. Descriptive statistics (means, SD, an SE for production pretests of two types of items by group) are displayed in Table 8.
Table 8. Means, SD, and SE for Grammar Production Pretests for Two Types of Item by Group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-I, +TA]</td>
<td>16</td>
<td>7.13</td>
<td>2.306</td>
<td>.576</td>
<td>.31</td>
<td>.793</td>
<td>.198</td>
</tr>
<tr>
<td>[-I, -TA]</td>
<td>15</td>
<td>5.60</td>
<td>2.473</td>
<td>.638</td>
<td>.27</td>
<td>.594</td>
<td>.153</td>
</tr>
<tr>
<td>[+I, -TA]</td>
<td>14</td>
<td>7.00</td>
<td>2.449</td>
<td>.655</td>
<td>.07</td>
<td>.267</td>
<td>.071</td>
</tr>
<tr>
<td>[++I, +TA]</td>
<td>14</td>
<td>7.57</td>
<td>2.027</td>
<td>.542</td>
<td>.64</td>
<td>1.336</td>
<td>.357</td>
</tr>
</tbody>
</table>

Note. Maximum score for Past Tense = 10. Maximum score for Present Subjunctive = 20.

The reactivity of verbal reports

In order to determine whether thinking aloud had an impact on the measures used in this study, data on task performance, text comprehension, production, recognition, and memory experiences were submitted to separate ANOVAs including think-aloud as the group factor.

Reactivity of think-aloud protocols on task performance

The reactivity of verbal reports on task performance was investigated by submitting the raw scores on the fill-in-the-blank task completed during reading by the minus incidental conditions to a one-way ANOVA. The results of this analysis indicated no statistically significant differences between the plus and minus think-aloud groups in their performance on lexical and grammatical items (Vocabulary: \( F(1, 30) = .051, p = .82 \); Present Subjunctive: \( F(1, 30) = .100, p = .75 \); Past tense: \( F(1, 30) = 1.720, p = .20 \)). Thus, thinking aloud during task performance did not appear to affect participants’ ability
to fill in the blanks of a text. Descriptive statistics (means, SD, an SE for performance on the fill-in task by think-aloud group) are displayed in Table 9.

Table 9. Means, SD, and SE for Fill-in Task Performance on Lexical and Grammatical Items by Think-aloud Group

<table>
<thead>
<tr>
<th>Group</th>
<th>Vocabulary</th>
<th>Present Subjunctive</th>
<th>Past Tense</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>[-I, +TA]</td>
<td>16</td>
<td>6.69</td>
<td>1.078</td>
</tr>
<tr>
<td>[-I, -TA]</td>
<td>15</td>
<td>6.80</td>
<td>1.656</td>
</tr>
</tbody>
</table>

Note. Maximum score for Vocabulary and Present Subjunctive = 10. Maximum score for Past Tense = 5.

Reactivity of think-aloud protocols on text comprehension

To investigate the reactivity of verbal reports on text comprehension, the scores for the plus versus minus think-aloud groups on the two reading comprehension post-exposure tests were submitted to separate ANOVAS. First, the scores obtained by the minus incidental groups (i.e., plus versus minus think-aloud groups) on the comprehension tests were submitted to a one-way ANOVA. No significant difference between groups was found in the comprehension of global or local ideas of the text containing target lexical items (Global Ideas: $F(1, 30) = .041, p = .84$; Local Ideas: $F(1, 30) = .007, p = .93$), or in the comprehension of the text containing target grammatical items ($F(1, 30) = 2.679, p = .11$). Second, the scores obtained by the plus incidental groups (i.e., plus versus minus think-aloud groups) on the comprehension tests were submitted to a one-way ANOVA. The results indicated no significant difference between groups in the comprehension of global or local ideas of the lexical item text (Global Ideas: $F(1, 27) = 1.800, p = .19$; Local Ideas: $F(1, 27) = 2.291, p = .14$) or in the
comprehension of the grammatical item text \( (F(1, 27) = 2.784, p = .10) \). Therefore, thinking aloud while reading did not appear to be reactive on participants’ reading comprehension in any of the conditions (i.e., plus versus minus incidental). Descriptive statistics (means, SD, and SE for performance on text comprehension by group) are displayed in Table 10.

**Table 10. Means, SD, and SE for Text Comprehension of Two Texts by Think-aloud Group**

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-Incidental</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[+TA]</td>
<td>16</td>
<td>5.31</td>
<td>2.938</td>
<td>.734</td>
<td>5.25</td>
<td>2.463</td>
<td>.616</td>
<td>6.19</td>
<td>2.994</td>
<td>.748</td>
</tr>
<tr>
<td>[-TA]</td>
<td>15</td>
<td>5.53</td>
<td>3.159</td>
<td>.816</td>
<td>5.33</td>
<td>3.109</td>
<td>.803</td>
<td>4.47</td>
<td>2.850</td>
<td>.733</td>
</tr>
<tr>
<td>+Incidental</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[+TA]</td>
<td>14</td>
<td>5.00</td>
<td>2.855</td>
<td>.763</td>
<td>4.36</td>
<td>3.003</td>
<td>.803</td>
<td>6.00</td>
<td>2.253</td>
<td>.602</td>
</tr>
<tr>
<td>[-TA]</td>
<td>14</td>
<td>6.50</td>
<td>3.057</td>
<td>.817</td>
<td>5.93</td>
<td>2.464</td>
<td>.659</td>
<td>7.43</td>
<td>2.277</td>
<td>.609</td>
</tr>
</tbody>
</table>

*Note. Maximum score = 10.*

**Reactivity of think-aloud protocols on memory experiences**

The reactivity of verbal reports on memory experiences was investigated by submitting the corrected scores on the remember-know task to separate repeated-measures ANOVAs with one between-subject variable (i.e., Group: think-aloud versus silent) and one within-subject variable with two levels (i.e., Memory Experience: remember versus know). The analyses on memory data from the minus incidental groups indicated no significant differences between groups (Vocabulary: \( F(1, 29) = .194, p = .66 \); Grammar: \( F(1, 29) = 2.385, p = .13 \)). The results of the analyses conducted on memory data from the plus incidental groups also support the lack of significant
difference between groups (Vocabulary: \( F(1, 26) = .004, p = .95 \); Grammar: \( F(1, 26) = .000, p = 1.00 \)). In other words, thinking aloud during task performance did not have a significant impact on participants’ subsequent experiences of recognition memory of the targeted items in either plus or minus incidental conditions.\(^{19}\) Descriptive statistics (means, SD, an SE for memory responses for two types of items by think-aloud group) are displayed in Table 11.

<table>
<thead>
<tr>
<th>Table 11. Means, SD, and SE for Reported Memory Experiences of Lexical and Grammatical Items by Think-aloud Group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linguistic Item</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Lexical Items</strong></td>
</tr>
<tr>
<td>[(-I, +TA)]</td>
</tr>
<tr>
<td>[(-I, -TA)]</td>
</tr>
<tr>
<td>[+I, +TA]</td>
</tr>
<tr>
<td>[+I, -TA]</td>
</tr>
<tr>
<td><strong>Grammatical Items</strong></td>
</tr>
<tr>
<td>[(-I, +TA)]</td>
</tr>
<tr>
<td>[(-I, -TA)]</td>
</tr>
<tr>
<td>[+I, +TA]</td>
</tr>
<tr>
<td>[+I, -TA]</td>
</tr>
</tbody>
</table>

Note. Maximum score = 10.

\(^{19}\) However, these results should be taken with caution due to the low power observed (Vocabulary: <.321; Grammar: <.08)
Reactivity of think-aloud protocols on recognition posttests

To investigate the reactivity of verbal reports on recognition of lexical and grammatical items, the raw scores on these tests were submitted to separate ANOVAs with one between-subject variable (i.e., Group: think-aloud versus silent) and one within-subject variable with three levels (i.e., Time: pretest versus immediate versus one-week-delayed). First, the results of the analysis conducted on the word meaning recognition scores obtained by the minus incidental groups (i.e., think-aloud versus silent) indicated no significant difference between groups ($F(1, 29) = 3.685, p = .06$) and no significant interaction Time x Group ($F(2, 58) = 2.482, p = .09$). In other words, thinking aloud while reading a glossed text plus completing a fill-in task did not have a significant impact on the ability to recognize the meaning of target and alternate words either immediately after exposure or one week later. Second, the results of the repeated-measures ANOVA run on the scores obtained by the minus incidental groups on the grammar recognition tests yielded no significant difference between groups ($F(1, 29) = 0.361, p = .55$) and no significant interaction Time x Group ($F(2, 58) = 0.250, p = .78$). Thinking aloud while reading a glossed text plus completing a fill-in task was not reactive on participants’ subsequent recognition of the target grammatical structure. The same pattern was found for the effects of thinking aloud in plus incidental conditions. The results of the repeated measures ANOVA on the word meaning recognition scores indicated no significant difference between groups ($F(1, 26) = 1.463, p = .23$) and no significant interaction Time x Group ($F(2, 52) = 1.063, p = .35$). Thinking aloud while reading a glossed text did not affect participants’ recognition of the target lexical items.
Similarly, the results of the repeated-measures ANOVA run on the grammar scores obtained by the plus incidental groups yielded no significant difference between groups ($F(1, 26) = .115, p = .73$) and no significant interaction Time x Group ($F(2, 52) = .291, p = .74$). In sum, thinking aloud while reading a glossed text did not significantly affect participants’ ability to recognize the target lexical and grammatical items immediately or one week after exposure.\(^{20}\) Descriptive statistics (means, SD, an SE for vocabulary and grammar recognition tests by think-aloud group) are displayed in Table 12.

Table 12. Means, SD, an SE for Word Meaning and Grammar Recognition by Time and Think-aloud Group

<table>
<thead>
<tr>
<th>Time</th>
<th>Think-aloud Group</th>
<th>Word Meaning</th>
<th>Grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Immediate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[-I, +TA]</td>
<td>16</td>
<td>7.44</td>
<td>5.329</td>
</tr>
<tr>
<td>[-I, -TA]</td>
<td>15</td>
<td>10.87</td>
<td>5.153</td>
</tr>
<tr>
<td>[+I, +TA]</td>
<td>14</td>
<td>3.64</td>
<td>1.906</td>
</tr>
<tr>
<td>[+I, -TA]</td>
<td>14</td>
<td>4.50</td>
<td>2.345</td>
</tr>
<tr>
<td>Delayed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[-I, +TA]</td>
<td>16</td>
<td>6.69</td>
<td>2.062</td>
</tr>
<tr>
<td>[-I, -TA]</td>
<td>15</td>
<td>9.53</td>
<td>3.167</td>
</tr>
<tr>
<td>[+I, +TA]</td>
<td>14</td>
<td>3.00</td>
<td>1.754</td>
</tr>
<tr>
<td>[+I, -TA]</td>
<td>14</td>
<td>3.93</td>
<td>2.235</td>
</tr>
</tbody>
</table>


\(^{20}\) Overall, the minus think-aloud groups outperformed the plus think-aloud groups in word meaning and grammar recognition and production tests although this difference was not significant. However, the power of the statistical analyses conducted turned out to be medium-low for word meaning production and recognition, and very low for grammar production and recognition (< 0.1). This result indicates that a significant difference between think-aloud and silent groups cannot be completely ruled out; yet this difference (if there were a significant difference) might be very small, and a much larger sample would be necessary to confirm the reactive effect of thinking aloud.
Reactivity of think-aloud protocols on written production posttests

To investigate the reactivity of verbal reports on production of lexical and grammatical items, the raw scores on these tests were submitted to separate repeated measures ANOVAs. First, the scores obtained by the minus incidental groups on the word meaning production tests were submitted to repeated measures ANOVA with one between-subject variable (i.e., Group: think-aloud versus silent) and one within-subject variable with three levels (i.e., Time: pretest versus immediate versus one-week-delayed). The results of this analysis indicated no significant difference between groups ($F(1, 29) = 3.575, p = .06$) and no significant interaction Time x Group ($F(2, 58) = 2.543, p = .08$). In other words, thinking aloud while reading a glossed text plus completing a fill-in task did not have a significant impact on the ability to translate target and alternate words either immediately one week after exposure. Then the scores obtained by the minus incidental groups on the grammar production tests were submitted to repeated measures ANOVA with one between-subject variable (i.e., Group: think-aloud versus silent) and one within-subject variable with three levels (i.e., Time: pretest versus immediate versus one-week-delayed). The results yielded no significant difference between groups ($F(1, 29) = .000, p = .99$) and no significant interaction Time x Group ($F(2, 58) = .007, p = .99$). Thus, thinking aloud during task performance did not appear to be reactive on participants’ subsequent production of old and new items of the target grammatical structure in the minus incidental condition. Similar results were found for reactivity of think-aloud protocols on participants’ production in the plus incidental conditions. The results of a repeated measures ANOVA conducted on the word meaning production
scores indicated no significant difference between groups \( F(1, 26) = 2.947, p = .09 \) and no significant interaction Time x Group \( F(2, 52) = 2.394, p = .10 \). The analysis conducted on grammar production scores yielded no significant difference between groups \( F(1, 26) = .376, p = .54 \) and no significant interaction Time x Group \( F(2, 52) = .465, p = .63 \). In sum, thinking aloud while reading a glossed text did not appear to significantly affect participants’ lexical and grammatical development in either minus or plus incidental conditions. Descriptive statistics (means, SD, an SE for vocabulary and grammar production tests by time and think-aloud group) are displayed in Table 13.

Table 13. Means, SD, an SE for Word Meaning and Grammar Production Tests by Time and Think-aloud Group

<table>
<thead>
<tr>
<th>Time</th>
<th>Think-aloud Group</th>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Word Meaning</td>
</tr>
<tr>
<td></td>
<td>( N )</td>
<td>( M )</td>
</tr>
<tr>
<td>Immediate</td>
<td>[-I, +TA]</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>[+I, -TA]</td>
<td>14</td>
</tr>
<tr>
<td>Delayed</td>
<td>[-I, +TA]</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>[+I, +TA]</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>[+I, -TA]</td>
<td>14</td>
</tr>
</tbody>
</table>

In sum, concurrent verbal reports in the present study did not appear to significantly affect task performance, text comprehension, reported memory experiences for lexical and grammatical items, word meaning and grammar production, or word meaning and grammar recognition.

**Time on task**

To determine whether the results could be attributable to differences on time spent to complete the experimental task (i.e., reading two texts in different glossing conditions), data on time on task were submitted to a one-way ANOVA. The results indicated a significant difference between groups in the time spent to read both the lexical item text ($F(4, 72) = 3.004, p = .02$) and the grammatical item text ($F(4, 72) = 5.393, p = .00$) with medium and large effect size, respectively (Lexical Item Text: Eta squared = .08; Grammatical Item Text: Eta squared = .18). A post-hoc Scheffé test showed that the [-incidental, +think-aloud] group spent significantly more time than the [+incidental, -think-aloud] group in reading both texts. Thinking aloud alone did not appear to slow down task completion since no significant difference was found between the [+think-aloud] and the [-think-aloud] groups either in the minus incidental or in the plus incidental conditions. Similarly, the fill-in task during reading did not seem to cause a delay in task completion since there was no significant difference between the [+incidental] and the [-incidental] groups in either the plus think-aloud or the minus think-aloud conditions. However, the results indicate that a combination of thinking aloud and completing a fill-in task while reading had a significant effect on time on task.
Descriptive statistics (means, SD, an SE for time on task by group) are displayed in Table 14.

Table 14. Means, SD, and SE for Time on Task (minutes) by Group and Text

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-I, +TA]</td>
<td>16</td>
<td>10.44</td>
<td>2.780</td>
<td>.695</td>
<td>10.31</td>
<td>3.459</td>
<td>.865</td>
</tr>
<tr>
<td>[-I, -TA]</td>
<td>15</td>
<td>8.47</td>
<td>1.885</td>
<td>.487</td>
<td>8.47</td>
<td>2.949</td>
<td>.761</td>
</tr>
<tr>
<td>[+I, +TA]</td>
<td>14</td>
<td>7.64</td>
<td>3.342</td>
<td>.893</td>
<td>7.14</td>
<td>3.159</td>
<td>.844</td>
</tr>
<tr>
<td>[+I, -TA]</td>
<td>14</td>
<td>7.14</td>
<td>2.282</td>
<td>.610</td>
<td>5.50</td>
<td>1.951</td>
<td>.522</td>
</tr>
<tr>
<td>[++I, +TA]</td>
<td>14</td>
<td>8.21</td>
<td>3.620</td>
<td>.967</td>
<td>7.86</td>
<td>2.878</td>
<td>.769</td>
</tr>
</tbody>
</table>

Since type of glossing is an independent variable in the present study, data were submitted to a one-way ANOVA with type of glossing as a factor to further confirm whether or not conditions differed statistically on time on task by type of glossing. In order to investigate type of glossing in the present study, think-aloud and silent conditions were combined so that the between-subject variable included three groups: [-I] (N = 31), [+I] (N = 28), and [++I], (N = 14). The results indicated a significant difference between groups in the time spent to read both the lexical item text \( (F(2, 72) = 3.929, p = .02) \) and the grammatical item text \( (F(2, 72) = 7.808, p = .00) \). A post-hoc Scheffé test showed that the minus incidental condition spent significantly more time than the plus incidental condition in reading both texts. However, none of the experimental conditions was significantly different from the control group. Descriptive statistics (means, SD, an SE for time on task by type of glossing) are displayed in Table 15.
Table 15. Means, SD, and SE for Time on Task (minutes) by Type of Glossing and Text

<table>
<thead>
<tr>
<th>Type of Glossing</th>
<th>Lexical Item Text</th>
<th>Grammatical Item Text</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
</tr>
<tr>
<td>[+I]</td>
<td>28</td>
<td>7.39</td>
</tr>
<tr>
<td>[++]I</td>
<td>14</td>
<td>8.21</td>
</tr>
</tbody>
</table>

Text effect

Two different texts were selected to include either target lexical items or target grammatical items. Thus, prior to investigating the effect of type of linguistic item on different dependent measures, it is necessary to minimize the possibility that differences between two item types are explained by differences between texts. For this reason, two texts of the same type and with similar length were selected. In addition, to ensure that the experimental texts did not lead to different levels of comprehension, data on global comprehension (i.e., comprehension of ideas that are not expressed by target unfamiliar items) were submitted to repeated measures ANOVA with one between-subject variable (Group) and one within-subject variable (Text). The results indicated no significant main effect for Text ($F(1, 68) = 3.105, p = .08$). Participants reached a similar level of comprehension of both texts regardless of the group. Therefore, any difference between lexical and grammatical items in participants’ memory experiences, awareness, and learning cannot be attributable to differences in text comprehension. Descriptive statistics (means, SD, and SE for global comprehension by text) are displayed in Table 16.
Table 16. Means, SD, and SE for Global Comprehension by Text

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-I, +TA]</td>
<td>16</td>
<td>5.31</td>
<td>2.938</td>
<td>.742</td>
<td>6.19</td>
<td>2.994</td>
<td>.641</td>
</tr>
<tr>
<td>[-I, -TA]</td>
<td>15</td>
<td>5.53</td>
<td>3.159</td>
<td>.767</td>
<td>4.47</td>
<td>2.850</td>
<td>.662</td>
</tr>
<tr>
<td>[+I, +TA]</td>
<td>14</td>
<td>5.00</td>
<td>2.855</td>
<td>.794</td>
<td>6.00</td>
<td>2.253</td>
<td>.685</td>
</tr>
<tr>
<td>[+I, -TA]</td>
<td>14</td>
<td>6.50</td>
<td>3.057</td>
<td>.794</td>
<td>5.50</td>
<td>2.277</td>
<td>.685</td>
</tr>
</tbody>
</table>

Note. Maximum score by text = 10.

Reliability

The reliability coefficients for the reading comprehension questionnaires and the production and recognition posttests were computed using Cronbach’s alpha. As shown in Table 17, all tests had high reliability.

Table 17. Tests Reliability

<table>
<thead>
<tr>
<th>Test</th>
<th>Cronbach’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text Comprehension Pretest</td>
<td>.69</td>
</tr>
<tr>
<td>Text Comprehension Test on the Lexical Item Text</td>
<td>.87</td>
</tr>
<tr>
<td>Text Comprehension Test on the Grammatical Item Text</td>
<td>.80</td>
</tr>
<tr>
<td>Word Meaning Recognition</td>
<td>.88</td>
</tr>
<tr>
<td>Word Meaning Production</td>
<td>.75</td>
</tr>
<tr>
<td>Grammar Recognition</td>
<td>.93</td>
</tr>
<tr>
<td>Grammar Production</td>
<td>.95</td>
</tr>
</tbody>
</table>
Part II. Experiences of Recognition Memory

Answer to Research Question #(1)

RQ#(1) Is there a relationship between reported experiences of recognition memory and:

(a) learners’ recognition of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

(b) learners’ written production of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

If so, will the relationship be maintained over a period of one week?

To answer RQ#(1a), a correlation analysis was conducted on reported memory experiences and recognition data. The results of this analysis indicated significant positive correlations between remembering and word meaning recognition in immediate and one-week-delayed posttests, and between remembering and grammar recognition in immediate and one-week-delayed posttests. All positive correlations were moderate. In addition, there was a weak negative correlation between guessing and grammar recognition in the immediate posttest. In sum, the experience of remembering appears to be related to high performance on recognition of both lexical and grammatical items, while the experience of knowing is not related to performance on recognition tests. Finally, the experience of guessing appears to be slightly related to low performance on immediate grammar recognition. Table 18 shows the results of the correlation analyses of memory experiences and recognition.
To answer RQ#(1b), data on reported memory experiences and production were submitted to a correlation analysis. The results of this analysis indicated significant positive correlations between remembering and word meaning production in immediate and one-week-delayed posttests, and between remembering and grammar production in immediate and one-week-delayed posttests. All positive correlations were moderate. In addition, there was a weak negative correlation between guessing and grammar production in the immediate posttest. In other words, the experience of remembering appears to be related to high performance on production of both lexical and grammatical items, while the experience of knowing is not related to performance on production tests. Finally, the experience of guessing appears to be slightly related to low performance on immediate grammar production. Table 19 shows the results of the correlation analyses.
Table 19. Correlation Analyses of Reported Memory Experiences and Production

<table>
<thead>
<tr>
<th>Production</th>
<th>Time</th>
<th>Reported Memory Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Immediate</td>
<td>Reported Memory Experience</td>
</tr>
<tr>
<td></td>
<td>Posttest</td>
<td>Remember</td>
</tr>
<tr>
<td>Lexical Items</td>
<td></td>
<td>$r = .630^{**}$</td>
</tr>
<tr>
<td>Immediate Posttest</td>
<td>$p = .000$</td>
<td>$p = .670$</td>
</tr>
<tr>
<td>Delayed Posttest</td>
<td>$r = .550^{**}$</td>
<td>$r = -.006$</td>
</tr>
<tr>
<td>$p = .000$</td>
<td>$p = .962$</td>
<td>$p = .208$</td>
</tr>
<tr>
<td>Grammatical Items</td>
<td>Delayed Posttest</td>
<td>$r = .559^{**}$</td>
</tr>
<tr>
<td>Immediate Posttest</td>
<td>$p = .000$</td>
<td>$p = .732$</td>
</tr>
<tr>
<td>Delayed Posttest</td>
<td>$r = .528^{**}$</td>
<td>$r = -.051$</td>
</tr>
<tr>
<td>$p = .000$</td>
<td>$p = .669$</td>
<td>$p = .071$</td>
</tr>
</tbody>
</table>

* The correlation is significant with the alpha level set at 0.05
** The correlation is significant with the alpha level set at 0.01

Part III. Levels of Awareness

Answer to Research Question #(2)

RQ#(2) Is there a relationship between the level of awareness reported during a reading comprehension task and the experience of recognition memory reported by learners for (i) unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

To answer RQ#(2), data on reported awareness and reported memory experiences were submitted to correlation analyses for both vocabulary and grammar. The results
indicate that noticing of lexical items (i.e., noticing of one or two word aspects)\textsuperscript{21} does not significantly correlate with remembering, knowing, or guessing experiences.

Similarly, the lack of verbal report for lexical items does not seem to correlate with any memory experience. With respect to grammar, the results indicate that understanding as measured by think-aloud protocols has a moderate positive correlation with remembering and a weak negative correlation with guessing. In other words, the more understanding reached by participants, the more remembering and the fewer guesses. The correlation analysis shows that noticing of grammatical items as measured by on-line procedures also positively correlates with remembering although to a lesser extent than understanding.\textsuperscript{22}

In contrast, the lack of verbal report is negatively correlated with remembering, although the correlation is weak. These results are summarized in Table 20.

\textsuperscript{21} Data on noticing of one word aspect and noticing of two word aspects have been submitted to correlation analyses. Because the results did not show any difference between types of noticing, both types have been conflated into a unique ‘noticing’ category for the sake of simplicity. A higher level of awareness for lexical items has not been considered in the analyses since only anecdotal evidence was found (see Process Measures section).

\textsuperscript{22} The same results were found for levels of reported awareness of grammatical items using a combination of on-line and off-line procedures. When no difference is found in relation to these measures, only the results for levels of reported awareness as measured by on-line procedures are reported.
Table 20. Correlation Analysis of Reported Levels of Awareness and Reported Memory Experiences

<table>
<thead>
<tr>
<th>Awareness</th>
<th>Memory Experience</th>
<th>Remember</th>
<th>Know</th>
<th>Guess</th>
</tr>
</thead>
</table>

| Lexical Items | | |
|---------------|----------------|---------|-----|-------|
| Noticing      | $r = -.153$     | $r = -.180$ | $r = .154$  |
|               | $p = .321$      | $p = .243$    | $p = .319$    |
| No verbal report | $r = -.067$    | $r = -.191$ | $r = .198$  |
|               | $p = .667$      | $p = .215$    | $p = .198$    |

| Grammatical Items | | |
|-------------------|----------------|---------|-----|-------|
| Understanding     | $r = .507^{**}$ | $r = .119$ | $r = -.367^{*}$ |
|                   | $p = .000$      | $p = .442$    | $p = .014$     |
| Noticing          | $r = .320^{*}$  | $r = -.224$ | $r = -.127$  |
|                   | $p = .034$      | $p = .144$    | $p = .410$     |
| No verbal report  | $r = -.320^{*}$ | $r = .224$   | $r = -.127$   |
|                   | $p = .034$      | $p = .144$    | $p = .410$     |

* The correlation is significant with the alpha level set at 0.05
** The correlation is significant with the alpha level set at 0.01

Answer to Research Question #(3)

RQ#(3) Is there a relationship between the level of awareness reported during a reading comprehension task and:

(a) learners’ recognition of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

(b) learners’ written production of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

If so, is this correlation maintained over a period of one week?

To answer RQ#(3a), data on reported awareness and scores on immediate and one-week-delayed recognition posttests were submitted to separate correlation analyses.
for vocabulary and grammar. The results indicate that noticing of lexical items has a strong positive correlation with immediate and one-week-delayed word meaning recognition, while the lack of verbal report does not correlate with vocabulary recognition. The results regarding grammar recognition\(^{23}\) show that understanding has a strong positive correlation with immediate grammar recognition, and a moderate correlation with one-week-delayed grammar recognition. Although to a lesser extent, noticing also significantly correlates with immediate and one-week-delayed grammar recognition, and lack of verbal report has a weak negative correlation with grammar recognition. This result indicates that higher levels of awareness are related to higher performance on grammar recognition tests, and that lack of verbal report is slightly related to lower performance on grammar recognition. These results are summarized in Table 21.

\(^{23}\) Data on recognition of old and new grammatical items have been submitted to correlation analyses. When the results do not show any difference between the two types of items, only the results for grammatical items overall are reported.
To answer RQ#(3b), data on reported awareness and scores on immediate and one-week-delayed production posttests were submitted to separate correlation analyses for vocabulary and grammar. The results indicate that noticing of lexical items has a moderate positive correlation with immediate and one-week delayed word meaning production. In addition, there is a weak negative correlation between no verbal report and immediate production; in other words, the lack of verbal reports appears to be slightly related to low performance on immediate word meaning production. The results regarding grammar production show that understanding as measured by think-aloud protocols has a strong positive correlation with immediate and one-week-delayed grammar production. Thus, the higher the level of awareness reached by participants, the higher performance on grammar production tests. Although to a lesser extent, noticing

<table>
<thead>
<tr>
<th>Linguistic Item</th>
<th>Awareness</th>
<th>Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Immediate</td>
<td>Delayed</td>
</tr>
<tr>
<td><strong>Lexical Items</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noticing</td>
<td>$r = .711^{**}$</td>
<td>$r = .712^{**}$</td>
</tr>
<tr>
<td></td>
<td>$p = .000$</td>
<td>$p = .000$</td>
</tr>
<tr>
<td>No verbal report</td>
<td>$r = .247$</td>
<td>$r = -.191$</td>
</tr>
<tr>
<td></td>
<td>$p = .106$</td>
<td>$p = .214$</td>
</tr>
<tr>
<td><strong>Grammatical Items</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understanding</td>
<td>$r = .760^{**}$</td>
<td>$r = .643^{**}$</td>
</tr>
<tr>
<td></td>
<td>$p = .000$</td>
<td>$p = .000$</td>
</tr>
<tr>
<td>Noticing</td>
<td>$r = .308^{*}$</td>
<td>$r = .437^{**}$</td>
</tr>
<tr>
<td></td>
<td>$p = .042$</td>
<td>$p = .003$</td>
</tr>
<tr>
<td>No verbal report</td>
<td>$r = -.308^{*}$</td>
<td>$r = -.437^{**}$</td>
</tr>
<tr>
<td></td>
<td>$p = .042$</td>
<td>$p = .003$</td>
</tr>
</tbody>
</table>

* The correlation is significant with the alpha level set at 0.05
** The correlation is significant with the alpha level set at 0.01
also significantly correlates with immediate and one-week-delayed grammar production. Finally, the lack of verbal reports has a weak negative correlation with grammar production. This result indicates that participants who produced fewer verbal reports than others tended to perform lower on the grammar production tests. Table 22 summarizes these results.

Table 22. Correlation Analysis of Reported Levels of Awareness and Production

<table>
<thead>
<tr>
<th>Linguistic Item</th>
<th>Awareness</th>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Immediate</td>
<td>Delayed</td>
</tr>
<tr>
<td><strong>Lexical Items</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noticing</td>
<td>( r = .590^{**} )</td>
<td>( r = .506^{**} )</td>
</tr>
<tr>
<td></td>
<td>( p = .000 )</td>
<td>( p = .000 )</td>
</tr>
<tr>
<td>No verbal report</td>
<td>( r = -.339^{*} )</td>
<td>( r = -.190 )</td>
</tr>
<tr>
<td></td>
<td>( p = .024 )</td>
<td>( p = .216 )</td>
</tr>
<tr>
<td><strong>Grammatical Items</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understanding</td>
<td>( r = .636^{**} )</td>
<td>( r = .741^{**} )</td>
</tr>
<tr>
<td></td>
<td>( p = .000 )</td>
<td>( p = .000 )</td>
</tr>
<tr>
<td>Noticing</td>
<td>( r = .343^{*} )</td>
<td>( r = .416^{**} )</td>
</tr>
<tr>
<td></td>
<td>( p = .023 )</td>
<td>( p = .005 )</td>
</tr>
<tr>
<td>No verbal report</td>
<td>( r = -.343^{*} )</td>
<td>( r = -.416^{**} )</td>
</tr>
<tr>
<td></td>
<td>( p = .023 )</td>
<td>( p = .005 )</td>
</tr>
</tbody>
</table>

* The correlation is significant with the alpha level set at 0.05  
** The correlation is significant with the alpha level set at 0.01

Part IV. Type of Glossing

Answer to Research Question #(4)

RQ#(4) Does type of glossing in a reading comprehension task have a differential effect on experiences of recognition memory reported by learners for (i) unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?
To answer RQ#(4), data on reported memory experiences were submitted to separate MANOVAs\textsuperscript{24} with one between-subject variable (i.e., type of glossing) and three dependent variables (i.e., remember, know, guess). To investigate whether type of glossing had a significant effect on reported memory experiences of lexical items, two MANOVAs were conducted, the first one including reports of target lexical items, and the second one including reports of both target and alternate lexical items (i.e., correct and incorrect options in the fill-task). The results of the first MANOVA yielded no significant effect for Group, $\Lambda = .899$, $F (2, 70) = 1.243$, $p = .288$. All groups reported similar amounts of remember, know and guess experiences for target lexical items. Similarly, the results of the second MANOVA indicated no significant main effect for Group ($\Lambda = .845$, $F (2, 70) = 1.986$, $p = .072$) in memory experiences for lexical items overall. Descriptive statistics (means, SD, and SE for reported memory experiences of lexical items by type of glossing) are displayed in Table 23.

\textsuperscript{24} A MANOVA was used to reduce the likelihood to make a Type I error.
Table 23. Means, SD, and SE for Reported Memory Experiences of Lexical Items by Type of Glossing

<table>
<thead>
<tr>
<th>Memory Experiences</th>
<th>Type of Glossing</th>
<th>Target Lexical Items</th>
<th>Lexical Items Overall (Target + Alternate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>M</td>
</tr>
<tr>
<td>Remembering</td>
<td>[-I]</td>
<td>31</td>
<td>2.52</td>
</tr>
<tr>
<td></td>
<td>[+I]</td>
<td>28</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>[++]I</td>
<td>14</td>
<td>2.36</td>
</tr>
<tr>
<td>Knowing</td>
<td>[-I]</td>
<td>31</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>[+I]</td>
<td>28</td>
<td>.79</td>
</tr>
<tr>
<td></td>
<td>[++]I</td>
<td>14</td>
<td>2.14</td>
</tr>
<tr>
<td>Guessing</td>
<td>[-I]</td>
<td>31</td>
<td>-1.13</td>
</tr>
<tr>
<td></td>
<td>[+I]</td>
<td>28</td>
<td>-.79</td>
</tr>
<tr>
<td></td>
<td>[++]I</td>
<td>14</td>
<td>-.79</td>
</tr>
</tbody>
</table>


Finally, the MANOVA conducted on memory experiences data obtained for target grammatical items indicated no significant main effect for Group, \( \Lambda = .883, F (2, 70) = 1.451, p = .200 \). The minus incidental group reported fewer guesses and remembered more target grammatical items than the other groups but this different was not significant. Descriptive statistics (means, SD, and SE for reported memory experiences of grammatical items by type of glossing) are displayed in Table 24.
Table 24. Mean, SD, and SE for Reported Memory Experiences of Grammatical Items by Type of Glossing

<table>
<thead>
<tr>
<th>Memory experiences</th>
<th>Type of Glossing</th>
<th>Target Grammatical Items</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remembering</td>
<td>[-I]</td>
<td></td>
<td>31</td>
<td>1.16</td>
<td>2.491</td>
<td>.401</td>
</tr>
<tr>
<td></td>
<td>[+I]</td>
<td></td>
<td>28</td>
<td>.64</td>
<td>2.215</td>
<td>.422</td>
</tr>
<tr>
<td></td>
<td>[++]</td>
<td></td>
<td>14</td>
<td>.07</td>
<td>1.542</td>
<td>.597</td>
</tr>
<tr>
<td>Knowing</td>
<td>[-I]</td>
<td></td>
<td>31</td>
<td>.55</td>
<td>2.173</td>
<td>.380</td>
</tr>
<tr>
<td></td>
<td>[+I]</td>
<td></td>
<td>28</td>
<td>.50</td>
<td>2.236</td>
<td>.400</td>
</tr>
<tr>
<td></td>
<td>[++]</td>
<td></td>
<td>14</td>
<td>-.71</td>
<td>1.684</td>
<td>.566</td>
</tr>
<tr>
<td>Guessing</td>
<td>[-I]</td>
<td></td>
<td>31</td>
<td>-.139</td>
<td>2.201</td>
<td>.345</td>
</tr>
<tr>
<td></td>
<td>[+I]</td>
<td></td>
<td>28</td>
<td>-.93</td>
<td>1.741</td>
<td>.363</td>
</tr>
<tr>
<td></td>
<td>[++]</td>
<td></td>
<td>14</td>
<td>.36</td>
<td>1.550</td>
<td>.514</td>
</tr>
</tbody>
</table>

Maximum score = 10

Answer to Research Question #(5)

RQ#(5) Does type of glossing in a reading comprehension task have a differential effect on levels of awareness reported by learners for (i) unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

To answer RQ#(5), data on reported awareness were submitted to a MANOVA with different levels of reported awareness as dependent variables, and type of glossing as a factor. To investigate whether type of glossing had a significant effect on reported noticing of lexical items, two MANOVAs were conducted: one including reports of target lexical items, and another one including reports of both target and alternate lexical items. The results of the first MANOVA yielded a significant effect for Group, $\Lambda = .284$, $F (2, 41) = 8.830$, $p = .000$, and the effect size was large (Eta squared = .467). The post-
hoc Scheffé tests indicated that the amount of reported noticing of target lexical items was significantly higher in the experimental groups than in the control group, while the amount of no verbal report was significantly higher in the control group. The experimental groups were not significantly different. The second MANOVA indicated a significant effect for Group, $\Lambda = .004, F (2, 41) = 143.878, p = .000$, with a large effect size (Eta squared = .938). The post-hoc Scheffé tests indicated that the amount of reported noticing of unfamiliar lexical items was significantly higher in the minus incidental group (which was exposed to target and alternate words) than in the plus incidental and control groups (which were exposed to target words only). In addition, the amount of reported noticing was significantly higher in the plus incidental group than in the control group. Descriptive statistics (means, SD, and SE for reported awareness of lexical items by type of glossing) are displayed in Table 25.

Table 25. Means, SD, and SE for Reported Awareness of Lexical Items

<table>
<thead>
<tr>
<th>Awareness</th>
<th>Type of Glossing</th>
<th>Target Lexical Items</th>
<th>Lexical Items Overall (Target + Alternate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$N$</td>
<td>$M$</td>
</tr>
<tr>
<td>Noticing</td>
<td>[-I, +TA]</td>
<td>16</td>
<td>6.19</td>
</tr>
<tr>
<td></td>
<td>[+I, +TA]</td>
<td>14</td>
<td>5.86</td>
</tr>
<tr>
<td></td>
<td>[+I, +TA]</td>
<td>14</td>
<td>2.71</td>
</tr>
<tr>
<td>No Verbal Report</td>
<td>[-I, +TA]</td>
<td>16</td>
<td>3.88</td>
</tr>
<tr>
<td></td>
<td>[+I, +TA]</td>
<td>14</td>
<td>7.29</td>
</tr>
</tbody>
</table>


The minus and plus incidental groups were not significantly different either in the amount of noticing reported overall, or in the amount of noticing of one versus two word aspects.
Finally, data on reported awareness of the target grammatical structure were submitted to two separate MANOVAs. One MANOVA with three dependent variables (i.e., understanding, noticing, and no verbal report) was conducted on data from the think-aloud groups (N = 44) where reported awareness was measured by concurrent verbal reports. Another MANOVA with three dependent variables (i.e., understanding, noticing, and low amount of verbal report) was conducted on data from all groups (N = 73) where reported awareness was measured by on-line and/or off-line measures (e.g., through a post-exposure questionnaire). The results of the first MANOVA yielded a significant effect for Group, \( \Lambda = .767, F(2, 41) = 2.835, p = .030 \), with a relatively large effect size (Eta squared = .124). The post-hoc Scheffé tests indicated that the amount of understanding was significantly higher in the minus incidental group than in the control group, while there was no significant difference between the minus and plus incidental groups, or between the plus incidental and control groups. Overall, the experimental groups reported more noticing than the control group, and the control group showed the highest amount of no verbal report. However, the amount of noticing and of no verbal report was not significantly different between groups. Descriptive statistics (means, SD, and SE for reported awareness of grammatical items as measured by online procedures) are displayed in Table 26.

---

26 As explained in the Process Measures section, participants were assigned one level of awareness of the grammatical structure (understanding versus noticing versus low amount of verbal report) based on the on-line and/or off-line measures. The combination of both types of measures allowed the researcher to conduct analyses of reported awareness on data from a larger sample of participants.
Table 26. Means, SD, and SE for Reported Awareness of Grammatical Items as Measured by Online Procedures

<table>
<thead>
<tr>
<th>Awareness as measured By online procedures</th>
<th>Type of Glossing</th>
<th>Target Grammatical Items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Understanding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[-I, +TA]</td>
<td>16</td>
<td>.31</td>
</tr>
<tr>
<td>[+I, +TA]</td>
<td>14</td>
<td>.07</td>
</tr>
<tr>
<td>[+I, +TA]</td>
<td>14</td>
<td>.00</td>
</tr>
<tr>
<td>Noticing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[-I, +TA]</td>
<td>16</td>
<td>5.50</td>
</tr>
<tr>
<td>[+I, +TA]</td>
<td>14</td>
<td>5.21</td>
</tr>
<tr>
<td>[+I, +TA]</td>
<td>14</td>
<td>2.86</td>
</tr>
<tr>
<td>No Verbal Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[-I, +TA]</td>
<td>16</td>
<td>4.50</td>
</tr>
<tr>
<td>[+I, +TA]</td>
<td>14</td>
<td>4.79</td>
</tr>
<tr>
<td>[+I, +TA]</td>
<td>14</td>
<td>7.14</td>
</tr>
</tbody>
</table>


The results of the second MANOVA yielded no significant effect for Group, Λ = .866, $F(2, 70) = 1.689, p = .128$, although the power of this analysis was medium (.627).

Overall, the minus incidental group showed higher understanding and noticing than the other groups but this difference was not significant. Descriptive statistics (means, SD, and SE for reported awareness of grammatical items as measured by online and/or offline procedures) are displayed in Table 27.
Table 27. Means, SD, and SE for Reported Awareness of Grammatical Items as Measured by Online and/or Offline Procedures

<table>
<thead>
<tr>
<th>Awareness as measured by online/offline procedures</th>
<th>Type of Glossing</th>
<th>Target Grammatical Items</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[-I]</td>
<td>31</td>
<td>.23</td>
<td>.425</td>
<td>.061</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[+I]</td>
<td>28</td>
<td>.11</td>
<td>.315</td>
<td>.064</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[++]</td>
<td>14</td>
<td>.00</td>
<td>.000</td>
<td>.091</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noticing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[-I]</td>
<td>31</td>
<td>.39</td>
<td>.495</td>
<td>.086</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[+I]</td>
<td>28</td>
<td>.29</td>
<td>.460</td>
<td>.090</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[++]</td>
<td>14</td>
<td>.29</td>
<td>.469</td>
<td>.127</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Amount of Verbal Report</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[-I]</td>
<td>31</td>
<td>.52</td>
<td>.508</td>
<td>.087</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[+I]</td>
<td>28</td>
<td>.71</td>
<td>.460</td>
<td>.091</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[++]</td>
<td>14</td>
<td>.71</td>
<td>.469</td>
<td>.129</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Maximum score = 1.

The difference between the results of these analyses comes from the fact that overall very few participants reported understanding of the grammatical structure. Thus, when considering both the plus think-aloud and the silent conditions (N = 73 versus N = 44), the difference between the minus incidental and control groups revealed by the analysis on online data turned out to be very small.

Answer to Research Question #(6)

RQ#(6) Does type of glossing in a reading comprehension task have a differential effect on:

(a) learners’ recognition of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?
(b) learners’ written production of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

If so, will the effect be maintained over a period of one week?

Effect of type of glossing on recognition

To answer the RQ#(6a), scores on word meaning and grammar recognition tests were submitted to separate repeated measures ANOVAs. Results on the effect of type of glossing on recognition of lexical items are presented first, followed by results regarding the effect of type of glossing recognition of grammatical items.

Effect of type of glossing on recognition of lexical items

To investigate the effect of type of glossing on recognition of lexical items, first an ANOVA with one between-subject variable (Type of Glossing) and one within-subject variable (Time) was conducted on scores on word meaning recognition of the target lexical items. The results of this analysis showed a significant main effect for Type of Glossing ($F(2, 140) = 10.551, p = .000$), a significant main effect for Time ($F(2, 140) = 85.605, p = .000$), and a significant interaction Time x Group ($F(2, 70) = 7.745, p = .000$). Results are shown in Table 28.

Table 28. ANOVA on Meaning Recognition of Target Lexical Items by Time and Group

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>2</td>
<td>352.356</td>
<td>176.178</td>
<td>85.605</td>
<td>.000*</td>
<td>.550</td>
<td>1.00</td>
</tr>
<tr>
<td>Group</td>
<td>2</td>
<td>40.362</td>
<td>20.181</td>
<td>10.551</td>
<td>.000*</td>
<td>.232</td>
<td>.986</td>
</tr>
<tr>
<td>Time * Group</td>
<td>4</td>
<td>63.358</td>
<td>15.939</td>
<td>7.745</td>
<td>.000*</td>
<td>.181</td>
<td>.997</td>
</tr>
</tbody>
</table>

*p < .05.
The post-hoc Scheffé tests indicated that (a) the experimental groups (i.e., plus and minus incidental) significantly outperformed the control group (i.e., the most incidental), and (b) the gain from pretest to immediate posttest was significantly higher in the experimental groups than in the control group. Descriptive statistics (means, SD, an SE for word meaning recognition of target lexical items by type of glossing and time) are displayed in Table 29.

Table 29. Means, SD, and SE for Meaning Recognition of Target Lexical Items by Type of Glossing and Time

<table>
<thead>
<tr>
<th>Type of Glossing</th>
<th>Pretest</th>
<th>Immediate</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>[I]</td>
<td>31</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>[+I]</td>
<td>28</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>[++I]</td>
<td>14</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note. Maximum score = 10

To further investigate the effect for time within each group, data on word meaning recognition of target lexical items were submitted to several One-Way ANOVAs with Time as a within-subject factor. Results for the minus incidental condition indicated a significant difference between pretest and immediate posttest ($F(1, 30) = 57.501, p = .000$), a significant difference between pretest and delayed posttest ($F(1, 30) = 69.227, p = .000$), and no significant difference between immediate posttest and delayed posttest ($F(1, 30) = 1.440, p = .213$). These results are displayed in Tables 30-32.
Table 30. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [-I] Group by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>252.016</td>
<td>252.016</td>
<td>57.501</td>
<td>.000*</td>
<td>.657</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p < .05.

Table 31. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [-I] Group by Time: Pretest versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>202.323</td>
<td>202.323</td>
<td>69.227</td>
<td>.000*</td>
<td>.698</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p < .05.

Table 32. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [-I] Group by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>2.726</td>
<td>2.726</td>
<td>1.440</td>
<td>.239</td>
<td>.046</td>
<td>.213</td>
</tr>
</tbody>
</table>

*p < .05.

Results for the plus incidental condition indicated a significant difference between pretest and immediate posttest ($F (1, 27) = 101.180, p = .000$), a significant difference between pretest and delayed posttest ($F (1, 27) = 81.765, p = .000$), and a significant difference between immediate posttest and delayed posttest ($F (1, 27) = 5.499, p = .027$).

These results are displayed in Tables 33-35.

Table 33. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [+I] Group by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>168.018</td>
<td>168.018</td>
<td>81.765</td>
<td>.000*</td>
<td>.752</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p < .05.
Table 34. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [+I] Group by Time: Pretest versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>202.323</td>
<td>202.323</td>
<td>69.227</td>
<td>.000*</td>
<td>.698</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p < .05.

Table 35. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [+I] Group by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>5.161</td>
<td>5.161</td>
<td>5.499</td>
<td>.027*</td>
<td>.169</td>
<td>.618</td>
</tr>
</tbody>
</table>

*p < .05.

Results for the most incidental condition (i.e., control group) indicated a significant difference between pretest and immediate posttest ($F(1, 13) = 23.400, p = .000$), a significant difference between pretest and delayed posttest ($F(1, 13) = 19.118, p = .001$), and no significant difference between immediate posttest and delayed posttest ($F(1, 13) = .807, p = .385$). These results are displayed in Tables 36-38.

Table 36. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [++] Group by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>5.143</td>
<td>5.143</td>
<td>23.400</td>
<td>.000*</td>
<td>.643</td>
<td>.994</td>
</tr>
</tbody>
</table>

*p < .05.

Table 37. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [++] Group by Time: Pretest versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>8.036</td>
<td>8.036</td>
<td>19.118</td>
<td>.001*</td>
<td>.595</td>
<td>.981</td>
</tr>
</tbody>
</table>

*p < .05.
Table 38. One-Way ANOVA on Meaning Recognition of Target Lexical Items within the [++I] Group by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>.321</td>
<td>.321</td>
<td>.807</td>
<td>.385</td>
<td>.058</td>
<td>.133</td>
</tr>
</tbody>
</table>

*\(p < .05\).

In sum, all groups significantly improved from pretest to immediate posttest, and maintained this gain one week later. The plus incidental group also performed significantly higher on the one-week-delayed posttest than on the pretest but experienced a significant loss from immediate to delayed posttest.

Finally, as shown by previous analyses, the plus and minus incidental groups were not significantly different in their ability to recognize the meaning of target words. In order to test whether the minus incidental group had recognized significantly more words overall (i.e. target and alternate words) than the plus incidental group (which had been exposed only to target words), the raw scores on the word meaning recognition tests were submitted to repeated measures ANOVA. The results of this analysis indicated a significant main effect for Time \((F(2, 140) = 78.187, p = .00)\), a significant main effect for Type of Glossing \((F(2, 70) = 29.481, p = .00)\), and a significant interaction Time x Type of Glossing \((F(4, 140) = 20.541, p = .00)\). A post-hoc Scheffé test revealed that the minus incidental group significantly outperformed both the plus incidental and the control groups on immediate and delayed posttests, and the plus incidental group significantly outperformed the control group. Descriptive statistics (means, SD, an SE for word meaning recognition of lexical items overall by type of glossing and time) are displayed in Table 39.
Table 39. Means, SD, and SE for Meaning Recognition of Lexical Items Overall by Type of Glossing and Time

<table>
<thead>
<tr>
<th>Type of Glossing</th>
<th>Pretest</th>
<th>Immediate</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>[-I]</td>
<td>31</td>
<td>.00 (.00)</td>
<td>9.10 (5.443)</td>
</tr>
<tr>
<td>[+I]</td>
<td>28</td>
<td>.00 (.00)</td>
<td>4.07 (2.142)</td>
</tr>
<tr>
<td>[++]I</td>
<td>14</td>
<td>.00 (.00)</td>
<td>.86 (.663)</td>
</tr>
</tbody>
</table>

Note. Maximum score = 20.

Figures 2 and 3 illustrate the results on word meaning recognition of target lexical items and overall, respectively.

Figure 2. Meaning Recognition of Target Lexical Items

Type of Glossing
- [- incidental]
- [+ incidental]
- [++] incidental]
Effect of type of glossing on recognition of grammatical items

To test whether type of glossing in a reading comprehension task had a differential effect on learners’ recognition of the target grammatical structure, scores on the grammar recognition tests were submitted to repeated measures ANOVA with one between-subject variable (Type of Glossing) and two within-subject variables (Time and Item Type, i.e., Old versus New). The results yielded no significant main effect for Type of Glossing ($F(1, 70) = 1.102, p = .338$), a significant main effect for Time ($F(2, 140) = 14.002, p = .000$), a significant main effect for Type of Item ($F(1, 70) = 10.8000, p = .002$), no significant interaction Time x Type of Glossing ($F(4, 140) = 1.547, p = .192$), no significant interaction Type of Item x Type of Glossing ($F(2, 70) = .946, p = .393$), no
significant interaction Time x Type of Item ($F (2, 140) = 1.862, p = .159$), and no
significant interaction Time x Type of Item x Type of Glossing ($F (4, 140) = .196, p =
.940$). Results are displayed in Table 40.

Table 40. ANOVA on Grammar Recognition by Time, Type of Item, and Group

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>2</td>
<td>194.428</td>
<td>97.214</td>
<td>14.002</td>
<td>.000*</td>
<td>.167</td>
<td>.998</td>
</tr>
<tr>
<td>Group</td>
<td>2</td>
<td>7.402</td>
<td>3.701</td>
<td>1.102</td>
<td>.338</td>
<td>.031</td>
<td>.236</td>
</tr>
<tr>
<td>Time * Group</td>
<td>4</td>
<td>42.962</td>
<td>10.741</td>
<td>1.547</td>
<td>.192</td>
<td>.042</td>
<td>.469</td>
</tr>
<tr>
<td>Item Type</td>
<td>1</td>
<td>7.583</td>
<td>7.583</td>
<td>10.800</td>
<td>.002*</td>
<td>.134</td>
<td>.900</td>
</tr>
<tr>
<td>Item type * Group</td>
<td>2</td>
<td>1.328</td>
<td>.664</td>
<td>.946</td>
<td>.393</td>
<td>.026</td>
<td>.208</td>
</tr>
<tr>
<td>Time * Item Type</td>
<td>2</td>
<td>3.147</td>
<td>1.574</td>
<td>1.862</td>
<td>.159</td>
<td>.026</td>
<td>.383</td>
</tr>
<tr>
<td>Time * Item Type * Group</td>
<td>4</td>
<td>.664</td>
<td>.166</td>
<td>.196</td>
<td>.940</td>
<td>.006</td>
<td>.091</td>
</tr>
</tbody>
</table>

*p < .05.

The results show that performance was significantly higher on old items than on
new items. The experimental groups (i.e., plus and minus incidental) outperformed the
control group but this difference was not significant. Descriptive statistics (means, SD, an
SE for grammar recognition by type of glossing, type of item, and time) are displayed in
Table 41.

Table 41. Means, SD, and SE for Grammar Recognition by Type of Glossing, Type of Item and Time

<table>
<thead>
<tr>
<th>Type of Glossing</th>
<th>Item Type</th>
<th>Pretest</th>
<th>Immediate</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>New .52</td>
<td>1.092</td>
<td>.175</td>
</tr>
<tr>
<td>[+I]</td>
<td>28</td>
<td>Old .64</td>
<td>.989</td>
<td>.168</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New .50</td>
<td>.839</td>
<td>.184</td>
</tr>
<tr>
<td>[++I]</td>
<td>14</td>
<td>Old .50</td>
<td>.650</td>
<td>.238</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New .57</td>
<td>.938</td>
<td>.260</td>
</tr>
</tbody>
</table>

Note. Maximum score for type of item = 10
To further investigate the effect for time within each group, data on grammar recognition were submitted to several One-Way ANOVAs with Time as a within-subject factor. Results for the minus incidental condition indicated a significant difference between pretest and immediate posttest \( (F(1, 30) = 12.991, p = .001) \), a significant difference between pretest and delayed posttest \( (F(1, 30) = 14.760, p = .001) \), and no significant difference between immediate posttest and delayed posttest \( (F(1, 30) = 2.598, p = .117) \). These results are displayed in Tables 42-44.

Table 42. One-Way ANOVA on Grammar Recognition within the [-I] Group by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>244.016</td>
<td>244.016</td>
<td>12.991</td>
<td>.001*</td>
<td>.302</td>
<td>.937</td>
</tr>
</tbody>
</table>

*p < .05.

Table 43. One-Way ANOVA on Grammar Recognition within the [-I] Group by Time: Pretest versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>382.516</td>
<td>382.516</td>
<td>14.760</td>
<td>.001*</td>
<td>.330</td>
<td>.960</td>
</tr>
</tbody>
</table>

*p < .05.

Table 44. One-Way ANOVA on Grammar Recognition within the [-I] Group by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>15.500</td>
<td>15.500</td>
<td>2.598</td>
<td>.117</td>
<td>.080</td>
<td>.345</td>
</tr>
</tbody>
</table>

*p < .05.

Results for the plus incidental condition indicated a significant difference between pretest and immediate posttest \( (F(1, 27) = 9.962, p = .004) \), a significant difference between pretest and delayed posttest \( (F(1, 27) = 6.738, p = .015) \), and a significant
difference between immediate posttest and delayed posttest \( (F(1, 27) = .060, p = .808) \).

These results are displayed in Tables 45-47.

Table 45. One-Way ANOVA on Grammar Recognition within the [+I] Group by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>111.446</td>
<td>111.446</td>
<td>9.962</td>
<td>.004*</td>
<td>.270</td>
<td>.860</td>
</tr>
</tbody>
</table>

*p < .05.

Table 46. One-Way ANOVA on Grammar Recognition within the [+I] Group by Time: Pretest versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>95.161</td>
<td>95.161</td>
<td>6.738</td>
<td>.015*</td>
<td>.200</td>
<td>.706</td>
</tr>
</tbody>
</table>

*p < .05.

Table 47. One-Way ANOVA on Grammar Recognition within the [+I] Group by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>.643</td>
<td>.643</td>
<td>.060</td>
<td>.808</td>
<td>.002</td>
<td>.056</td>
</tr>
</tbody>
</table>

*p < .05.

Results for the most incidental condition (i.e, control group) indicated no significant difference between pretest and immediate posttest \( (F(1, 13) = 1.087, p = .316) \), no significant difference between pretest and delayed posttest \( (F(1, 13) = 3.020, p = .106) \), and no significant difference between immediate posttest and delayed posttest \( (F(1, 13) = 1.407, p = .257) \). These results are displayed in Tables 48-50.

Table 48. One-Way ANOVA on Grammar Recognition within the [++I] Group by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>2.893</td>
<td>2.893</td>
<td>1.087</td>
<td>.316</td>
<td>.077</td>
<td>.162</td>
</tr>
</tbody>
</table>

*p < .05.
Table 49. One-Way ANOVA on Grammar Recognition within the [++] Group by Time: Pretest versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>41.286</td>
<td>41.286</td>
<td>3.020</td>
<td>.106</td>
<td>.189</td>
<td>.363</td>
</tr>
</tbody>
</table>

*p < .05.

Table 50. One-Way ANOVA on Grammar Recognition within the [++] Group by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>22.321</td>
<td>22.321</td>
<td>1.407</td>
<td>.257</td>
<td>.098</td>
<td>.196</td>
</tr>
</tbody>
</table>

*p < .05.

In sum, these results show that the experimental groups significantly improved from pretest to immediate posttest, and retained this gain over a period of a week, while the control group did not significantly change over time. However, the experimental groups were not significantly different from the control group. Figures 4 and 5 illustrate the results on grammar recognition of old and new items, respectively.
Figure 4. Grammar Recognition: Old Items

Figure 5. Grammar Recognition: New Items

Type of Glossing
▲ [- incidental]
● [+ incidental]
■ [++ incidental]

201
Effect of type of glossing on written production

To answer the RQ#(6b), scores on word meaning and grammar production tests were submitted to separate repeated measures ANOVAs. Results on the effect of type of glossing on meaning production of lexical items are presented first, followed by results regarding the effect of type of glossing production of grammatical items.

Effect of type of glossing on written production of the meaning of lexical items

To investigate the effect of type of glossing on meaning production of lexical items, scores on word meaning production of the target words were submitted to repeated measures ANOVA with one between-subject variable (Type of Glossing) and one within-subject variable (Time). The results of this analysis yielded a significant main effect for Type of Glossing ($F(2, 70) = 6.651, p = .002$), a significant main effect for Time, ($F(2, 140) = 37.686, p = .000$), and a significant interaction Time x Type of Glossing ($F(2, 70) = 5.071, p = .001$). As shown in Table 51, for all effects a large effect size and high power were found.

Table 51. ANOVA on Meaning Production of Target Lexical Items by Time and Group

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>2</td>
<td>53.739</td>
<td>26.870</td>
<td>37.686</td>
<td>.000*</td>
<td>.350</td>
<td>1.00</td>
</tr>
<tr>
<td>Group</td>
<td>2</td>
<td>35.209</td>
<td>35.209</td>
<td>6.651</td>
<td>.002*</td>
<td>.160</td>
<td>.902</td>
</tr>
<tr>
<td>Time * Group</td>
<td>4</td>
<td>14.464</td>
<td>3.616</td>
<td>5.071</td>
<td>.001*</td>
<td>.127</td>
<td>.961</td>
</tr>
</tbody>
</table>

*p < .05.

The post-hoc Scheffé tests showed that (a) the experimental groups (i.e., plus and minus incidental) significantly outperformed the control group (i.e., the most incidental) overall, and (b) the gain from pretest to immediate posttest was significantly higher in the
experimental groups than in the control group. Descriptive statistics (means, SD, an SE for meaning production of target lexical items by type of glossing and time) are displayed in Table 52.

Table 52. Means, SD, and SE for Meaning Production of Target Lexical Items by Type of Glossing and Time

<table>
<thead>
<tr>
<th>Type of Glossing</th>
<th>Pretest</th>
<th>Immediate</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-I]</td>
<td>N = 31</td>
<td>M = .00</td>
<td>SD = .00</td>
</tr>
<tr>
<td>[+I]</td>
<td>N = 28</td>
<td>M = .00</td>
<td>SD = .00</td>
</tr>
<tr>
<td>[++I]</td>
<td>N = 14</td>
<td>M = .00</td>
<td>SD = .00</td>
</tr>
</tbody>
</table>

Note. Maximum score = 10.

To investigate the effect of time within each group, data on word meaning production of target lexical items were submitted to several One-Way ANOVAs with Time as a within-subject factor. Results for the minus incidental condition indicated a significant difference between pretest and immediate posttest (\(F(1, 30) = 26.224, p = .000\)), a significant difference between pretest and delayed posttest (\(F(1, 30) = 45.345, p = .000\)), and no significant difference between immediate posttest and delayed posttest (\(F(1, 30) = 1.825, p = .187\)). These results are displayed in Tables 53-55.

Table 53. One-Way ANOVA on Meaning Production of Target Lexical Items within the [-I] Group by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>28.452</td>
<td>28.452</td>
<td>26.224</td>
<td>.000*</td>
<td>.466</td>
<td>.999</td>
</tr>
</tbody>
</table>

*p < .05.
Table 54. One-Way ANOVA on Meaning Production of Target Lexical Items within the [-I] Group by Time: Pretest versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>40.323</td>
<td>40.323</td>
<td>45.345</td>
<td>.000*</td>
<td>.602</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p < .05.

Table 55. One-Way ANOVA on Meaning Production of Target Lexical Items within the [-I] Group by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>1.032</td>
<td>1.032</td>
<td>1.825</td>
<td>.187</td>
<td>.057</td>
<td>.258</td>
</tr>
</tbody>
</table>

*p < .05.

Results for the plus incidental condition indicated a significant difference between pretest and immediate posttest \( (F (1, 27) = 28.972, p = .000) \), a significant difference between pretest and delayed posttest \( (F (1, 27) = 40.114, p = .000) \), and no significant difference between immediate posttest and delayed posttest \( (F (1, 27) = .051, p = .823) \).

These results are displayed in Tables 56-58.

Table 56. One-Way ANOVA on Meaning Production of Target Lexical Items within the [+I] Group by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>37.786</td>
<td>37.786</td>
<td>28.972</td>
<td>.000*</td>
<td>.518</td>
<td>.999</td>
</tr>
</tbody>
</table>

*p < .05.

Table 57. One-Way ANOVA on Meaning Production of Target Lexical Items within the [+I] Group by Time: Pretest versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>36.161</td>
<td>36.161</td>
<td>40.114</td>
<td>.000*</td>
<td>.598</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p < .05.
Table 58. One-Way ANOVA on Meaning Production of Target Lexical Items within the [+I] Group by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>.018</td>
<td>.018</td>
<td>.051</td>
<td>.823</td>
<td>.002</td>
<td>.055</td>
</tr>
</tbody>
</table>

*p < .05.

Results for the most incidental condition (i.e., control group) indicated no significant difference between pretest and immediate posttest ($F (1, 13) = 1.000, p = .336$), no significant difference between pretest and delayed posttest ($F (1, 13) = 4.452, p = .055$), and a significant difference between immediate posttest and delayed posttest ($F (1, 13) = 5.200, p = .040$). These results are displayed in Tables 59-61.

Table 59. One-Way ANOVA on Meaning Production of Target Lexical Items within the [++I] Group by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>.036</td>
<td>.036</td>
<td>1.000</td>
<td>.336</td>
<td>.071</td>
<td>.153</td>
</tr>
</tbody>
</table>

*p < .05.

Table 60. One-Way ANOVA on Meaning Production of Target Lexical Items within the [++I] Group by Time: Pretest versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>.893</td>
<td>.893</td>
<td>4.452</td>
<td>.055</td>
<td>.255</td>
<td>.497</td>
</tr>
</tbody>
</table>

*p < .05.

Table 61. One-Way ANOVA on Meaning Production of Target Lexical Items within the [++I] Group by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>.571</td>
<td>.571</td>
<td>5.200</td>
<td>.040*</td>
<td>.286</td>
<td>.560</td>
</tr>
</tbody>
</table>

*p < .05.
The results regarding the effect of time on meaning production reveal that the experimental groups significantly improved from pretest to immediate posttest, and retained this gain over a period of a week. The control group did not improve from pretest to immediate and delayed posttests but showed a significant gain from immediate posttest to one-week-delayed posttest. This significant change in the control group may be explained by a practice effect, as a result of completing meaning production and recognition tests during the second session of the experiment. Overall, the experimental groups performed significantly higher than the control group.

As shown in Figure 6, the plus and minus incidental groups were not significantly different in their ability to translate target words.

Figure 6. Meaning Production of Target Lexical Items
In order to test whether the minus incidental group had translated significantly more words overall (i.e., target and alternate words in the fill-in task) than the plus incidental group, the raw scores on the vocabulary production tests were submitted to repeated measures ANOVA. The results of this analysis indicated a significant main effect for Type of Glossing \( F (2, 70) = 8.532, p = .00 \), a significant main effect for Time \( F (2, 140) = 29.904, p = .00 \), and a significant interaction Time x Type of Glossing \( F (4, 140) = 6.478, p = .00 \). Descriptive statistics (means, SD, and SE for meaning production of lexical items overall by type of glossing) are displayed in Table 62. As the means indicate, the minus incidental group outperformed both the plus incidental and the control groups on immediate and delayed posttests. However, a post-hoc Scheffé test revealed that the minus incidental group was significantly different from the control group but not from the plus incidental group in both immediate and one-week-delayed posttests.

Table 62. Means, SD, and SE for Meaning Production of Lexical Items Overall by Type of Glossing and Time

<table>
<thead>
<tr>
<th>Type of Glossing</th>
<th>Pretest</th>
<th>Immediate</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( N )</td>
<td>( M )</td>
<td>( SD )</td>
</tr>
<tr>
<td>([-I])</td>
<td>31</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>([+I])</td>
<td>28</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>([++I])</td>
<td>14</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

Note. Maximum score = 20.
To investigate whether type of glossing in a reading comprehension task had a
differential effect on learners’ production of the target grammatical structure, scores on
the grammar production tests were submitted to repeated measures ANOVA with one
between-subject variable (Type of Glossing) and two within-subject variables (Time and
Type of Item, i.e., Old versus New). Results yielded no significant main effect for Type
of Glossing ($F (2, 70) = 1.102, p = .338$), a significant main effect for Time ($F (2, 140) =
11.104, p = .000$), a significant main effect for Type of Item ($F (1, 70) = 12.656, p =
.001$), no significant interaction Time x Type of Glossing ($F (4, 140) = 1.091, p = .364$),
no significant interaction Type of Item x Type of Glossing ($F (2, 70) = .839, p = .437$), a
significant interaction Time x Type of Item ($F (2, 140) = 4.277, p = .016$), and a
significant triple interaction Type x Type of Item x Type of Glossing ($F (4, 140) = 3.620,
p = .008$). Results are shown in Table 63.

The post-hoc Scheffé tests showed that (a) overall, performance was significantly
higher on old items than on new items, (b) the gain from pretest to immediate posttest
was significantly higher for old items than for new items, and (c) the gain of old items from pretest to immediate posttest was significantly higher in the experimental groups (i.e., plus and minus incidental) than in the control group. However, the main effect for Group (i.e., Type of Glossing) was not significant. The means shown in Table 64 indicate that the control group performed higher on the delayed posttest than on the immediate posttest, which reduced the difference between the experimental and the control conditions.

Table 64. Means, SD, and SE for Grammar Production by Type of Glossing, Type of Item and Time

<table>
<thead>
<tr>
<th>Type of Glossing</th>
<th>Item Type</th>
<th>Pretest</th>
<th>Immediate</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-I]</td>
<td>Old</td>
<td>.19</td>
<td>.477</td>
<td>.069</td>
</tr>
<tr>
<td></td>
<td>New</td>
<td>.10</td>
<td>.301</td>
<td>.050</td>
</tr>
<tr>
<td>[+I]</td>
<td>Old</td>
<td>.07</td>
<td>.262</td>
<td>.073</td>
</tr>
<tr>
<td></td>
<td>New</td>
<td>.04</td>
<td>.189</td>
<td>.052</td>
</tr>
<tr>
<td>[++I]</td>
<td>Old</td>
<td>.14</td>
<td>.363</td>
<td>.103</td>
</tr>
<tr>
<td></td>
<td>New</td>
<td>.14</td>
<td>.363</td>
<td>.074</td>
</tr>
</tbody>
</table>

Note. Maximum score by type of item = 10.

To further investigate the effect for time within each group, data on grammar production were submitted to several One-Way ANOVAs with Time as a within-subject factor. Results for the minus incidental condition indicated a significant difference between pretest and immediate posttest \((F (1, 30) = 7.614, p = .010)\), a significant difference between pretest and delayed posttest \((F (1, 30) = 9.026, p = .005)\), and no significant difference between immediate posttest and delayed posttest \((F (1, 30) = .620, p = .437)\). These results are displayed in Tables 65-67.
Table 65. One-Way ANOVA on Grammar Production within the [-I] Group by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>167.806</td>
<td>167.806</td>
<td>7.614</td>
<td>.010*</td>
<td>.202</td>
<td>.761</td>
</tr>
</tbody>
</table>

*p < .05.

Table 66. One-Way ANOVA on Grammar Production within the [-I] Group by Time: Pretest versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>224.581</td>
<td>224.581</td>
<td>9.026</td>
<td>.005*</td>
<td>.231</td>
<td>.828</td>
</tr>
</tbody>
</table>

*p < .05.

Table 67. One-Way ANOVA on Grammar Production within the [-I] Group by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>4.129</td>
<td>4.129</td>
<td>.620</td>
<td>.437</td>
<td>.020</td>
<td>.119</td>
</tr>
</tbody>
</table>

*p < .05.

Results for the plus incidental condition indicated a significant difference between pretest and immediate posttest ($F(1, 27) = 8.850, p = .006$), a significant difference between pretest and delayed posttest ($F(1, 27) = 9.403, p = .005$), and no significant difference between immediate posttest and delayed posttest ($F(1, 27) = .039, p = .844$).

These results are displayed in Tables 68-70.

Table 68. One-Way ANOVA on Grammar Production within the [+I] Group by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>135.161</td>
<td>135.161</td>
<td>8.850</td>
<td>.006*</td>
<td>.247</td>
<td>.818</td>
</tr>
</tbody>
</table>

*p < .05.
Results for the most incidental condition (i.e., control group) indicated no significant difference between pretest and immediate posttest \( (F(1, 13) = .089, p = .770) \), no significant difference between pretest and delayed posttest \( (F(1, 13) = 1.079, p = .318) \), and no significant difference between immediate posttest and delayed posttest \( (F(1, 13) = 1.443, p = .251) \). These results are displayed in Tables 71-73.
Table 73. One-Way ANOVA on Grammar Production within the [++] Group by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>14.286</td>
<td>14.286</td>
<td>1.443</td>
<td>.251</td>
<td>.100</td>
<td>.200</td>
</tr>
</tbody>
</table>

*p < .05.

To summarize, these results show that the experimental groups significantly improved from pretest to immediate posttest, and retained this gain over a period of a week, while the control group did not significantly change over time. However, the experimental groups were not significantly different from the control group. Figures 7 and 8 illustrate the results on grammar production of old and new items, respectively.

Figure 7. Grammar Production: Old Items
Answer to Research Question #(7)

RQ#(7) Does type of glossing in a reading comprehension task have a differential effect on learners’ text comprehension?

To answer RQ#(7), scores on the two text comprehension tests were submitted to separate ANOVAs. A repeated measures ANOVA with one between-subject variable (Type of Glossing) and one within-subject variable (Idea Type: Local versus Global) was conducted on comprehension scores for the text containing target lexical items (i.e., lexical item text). Table 74 shows that the results of this analysis yielded no main effect
for Type of Glossing ($F(2, 70) = 2.879, p = .063$), a significant main effect for Type of Idea ($F(1, 70) = 45.896, p = .000$), and a significant interaction Type of Idea x Type of Glossing ($F(2, 70) = 23.920, p = .000$). Large effect size and high power were found for the significant effects.

Table 74. ANOVA for Text Comprehension by Type of Idea and Group on the Lexical Item Text

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Idea</td>
<td>1</td>
<td>115.097</td>
<td>115.097</td>
<td>45.896</td>
<td>.000*</td>
<td>.396</td>
<td>1.00</td>
</tr>
<tr>
<td>Group</td>
<td>2</td>
<td>36.425</td>
<td>18.213</td>
<td>2.879</td>
<td>.063</td>
<td></td>
<td>.546</td>
</tr>
<tr>
<td>Type of Idea * Group</td>
<td>4</td>
<td>119.975</td>
<td>59.988</td>
<td>23.920</td>
<td>.000*</td>
<td>.406</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p < .05.

The post-hoc Scheffé tests showed that (a) performance was significantly higher on comprehension of global ideas (i.e., ideas not expressed by the target items) than on local ideas (i.e., ideas expressed by the target items), and (b) the experimental groups (i.e., plus and minus incidental) significantly outperformed the control group on local comprehension. Descriptive statistics (means, SD, and SE for text comprehension by type of glossing and type of idea on the lexical item text) are displayed in Table 75.

Table 75. Means, SD, and SE for Text Comprehension on the Lexical Item Text by Group and Type of Idea

<table>
<thead>
<tr>
<th>Type of Idea</th>
<th>Group</th>
<th>N</th>
<th>Text Comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>M</td>
</tr>
<tr>
<td><strong>Global</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[-I]</td>
<td>31</td>
<td>5.42</td>
<td>2.997</td>
</tr>
<tr>
<td>[+I]</td>
<td>28</td>
<td>5.75</td>
<td>3.002</td>
</tr>
<tr>
<td>[++]</td>
<td>14</td>
<td>6.07</td>
<td>2.814</td>
</tr>
<tr>
<td><strong>Local</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[-I]</td>
<td>31</td>
<td>5.29</td>
<td>2.747</td>
</tr>
<tr>
<td>[+I]</td>
<td>28</td>
<td>5.14</td>
<td>2.812</td>
</tr>
<tr>
<td>[++]</td>
<td>14</td>
<td>1.14</td>
<td>.663</td>
</tr>
</tbody>
</table>

Note. Maximum score by type of idea = 10.
Figure 9 illustrates the results obtained for text comprehension by type of glossing and type of idea on the lexical item text.

Finally, scores on global comprehension of the grammatical item text\(^{27}\) were submitted to a one-way ANOVA. As shown in Table 76, results yielded no significant main effect for Type of Glossing ($F(2, 72) = 2.232, p = .115$).

Table 76. ANOVA for Text Comprehension by Type of Idea and Group on the Grammatical Item Text

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>2</td>
<td>30.884</td>
<td>15.442</td>
<td>2.232</td>
<td>.115</td>
</tr>
</tbody>
</table>

\(^{27}\) Because the meaning of the targeted verbs was familiar to participants, local comprehension of the grammatical item text was not addressed.
Descriptive statistics (means, SD, an SE for text comprehension by type of glossing on the grammatical item text) are displayed in Table 77.

<table>
<thead>
<tr>
<th>Type of Glossing</th>
<th>Text Comprehension of Global Ideas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>[-I]</td>
<td>31</td>
</tr>
<tr>
<td>[+I]</td>
<td>28</td>
</tr>
<tr>
<td>[++]</td>
<td>14</td>
</tr>
</tbody>
</table>

Note. Maximum score = 10

Part IV. Type of Linguistic Item

Answer to Research Question #(8)

RQ#(8) Does type of linguistic item (i.e., lexical versus grammatical) in a reading comprehension task have a differential effect on experiences of recognition memory reported by learners?

To answer the RQ#(8), data on reported memory experiences were submitted to a MANOVA with three dependent variables (i.e., remembering, knowing, guessing) and type of linguistic item as a factor. The results of this analysis indicated a significant effect for Type of Linguistic Item, $\Lambda = .731$, $F(1, 144) = 17.418$, $p = .000$, and the effect size was large (Eta squared = .269). Descriptive statistics (means, SD, and SE for reported memory experiences by type of linguistic item) are displayed in Table 78.
Table 78. Means, SD, and SE for Reported Memory Experiences by Type of Linguistic Item

<table>
<thead>
<tr>
<th>Type of Linguistic Item</th>
<th>Remembering</th>
<th></th>
<th></th>
<th>Knowing</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Guessing</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>SE</td>
<td>M</td>
<td>SD</td>
<td>SE</td>
<td>M</td>
<td>SD</td>
<td>SE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexical Items</td>
<td>73</td>
<td>2.30</td>
<td>2.093</td>
<td>.254</td>
<td>1.25</td>
<td>2.397</td>
<td>.266</td>
<td>-.51</td>
<td>1.733</td>
<td>.219</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grammatical Items</td>
<td>73</td>
<td>.75</td>
<td>2.241</td>
<td>.254</td>
<td>.29</td>
<td>2.144</td>
<td>.266</td>
<td>-.88</td>
<td>2.000</td>
<td>.219</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Maximum score = 10.

The means in Table 78 show that lexical items received more remember and know responses than grammatical items. The analysis indicated that the difference in remembering was significant ($F (1, 144) = 18.605, p = .000$) as well as the difference in knowing ($F (1, 144) = 6.491, p = .012$). The difference in guessing, instead, was not significant ($F (1, 144) = 1.426, p = .234$). Figure 10 illustrates this pattern.
Answer to Research Question #(9)

RQ#(9) Does type of linguistic item (i.e., lexical versus grammatical) in a reading comprehension task have a differential effect on levels of awareness reported by learners?

To answer the RQ#(9), data on reported levels of awareness were submitted to two different analyses. First, instances of noticing and instances of no verbal report for lexical and grammatical items as measured by think-aloud protocols were submitted to a MANOVA with one between-subject variable (i.e., type of linguistic item) and two dependent variables (i.e., noticing versus no verbal report). The results of this analysis indicated no significant effect for Type of Linguistic Item, $\Lambda = .985, F(1, 86) = .667, p = .516$. Descriptive statistics (means, SD, and SE for reported awareness by type of linguistic item) are displayed in Table 79.

<table>
<thead>
<tr>
<th>Type of Linguistic Item</th>
<th>Noticing</th>
<th>No Verbal Report</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>$M$</td>
</tr>
<tr>
<td>Lexical Items</td>
<td>44</td>
<td>4.98</td>
</tr>
<tr>
<td>Grammatical Items</td>
<td>44</td>
<td>4.57</td>
</tr>
</tbody>
</table>

Note. Maximum score = 10.

Second, data on reported awareness of lexical items were submitted to a MANOVA with one between-subject variable (i.e., Level of Awareness of Grammatical Items, as measured by on-line and/or off-line procedures: Understanding versus Noticing versus Low Amount of Verbal Report) and two dependent variables (Noticing of Lexical

28 Understanding is not included as dependent variable since only one case of awareness at a high level was found for lexical items.
Items versus No Verbal Report of Lexical Items). The results of this analysis indicated a significant effect for Type of Linguistic Item, $\Lambda = .737, F(2, 41) = 3.296, p = .015$, and the effect size was large (Eta squared = .141). Descriptive statistics (mean, SD, and SE for reported awareness by type of linguistic item) are displayed in Table 80 below. The means indicate that the group showing awareness of the grammatical structure at the highest level (i.e., the Understanding group) reported more noticing of lexical items than the Noticing group, which in turn reported more noticing of lexical items than the Low Amount of Verbal Report group. In addition, the Low Amount of Verbal Report group showed the highest amount of no verbal report of lexical items, while the Understanding group showed the lowest amount of no verbal report of lexical items.

Table 80. Reported Awareness of Lexical Items by Reported Level of Awareness of Grammatical Items

<table>
<thead>
<tr>
<th>Level of Awareness of Grammatical Items</th>
<th>Noticing of Lexical Items</th>
<th>No Verbal Report of Lexical Items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
</tr>
<tr>
<td>Understanding</td>
<td>6</td>
<td>7.50</td>
</tr>
<tr>
<td>Noticing</td>
<td>12</td>
<td>5.92</td>
</tr>
<tr>
<td>Low Amount of Verbal Report</td>
<td>26</td>
<td>3.96</td>
</tr>
</tbody>
</table>

Note. Maximum score = 10.

A post-hoc Scheffé test revealed that the Understanding and Low Amount of Verbal Report groups were significantly different in amount of noticing and amount of no verbal report of lexical items, while the Noticing group was not significantly different from any group. This pattern is illustrated in Figure 11. These results should be taken with caution, however, due to the low number of participants in the Understanding group.
In sum, the results indicate that type of linguistic item in this study, that is, vocabulary versus grammar, did not have a significant impact on reported awareness of target items. Although the level and the amount of reported awareness overall varied across individuals, participants appeared to report similar amount of awareness for grammatical and lexical items.

**Answer to Research Question #(10)**

RQ#(10) Does type of linguistic item (i.e., lexical versus grammatical) in a reading comprehension task have a differential effect on learners’ recognition and written production? If so, will the effect be maintained over a period of one week?

To answer RQ#(10) the percentage of correct answers on vocabulary and grammar recognition and production tests were submitted to separate repeated measures
ANOVA with two within-subject variables (i.e., Type of Linguistic Item and Time).

Results on recognition are presented first, followed by results on production.

**Effect of type of linguistic item on recognition**

The results of the ANOVA for recognition yielded no significant main effect for Type of Linguistic Item ($F(1, 72) = 2.121, p = .150$), a significant main effect for Time ($F(2, 144) = 12.722, p = .000$), and no significant interaction Type of Linguistic Item x Time ($F(2, 144) = .592, p = .555$). Table 81 shows these results.

Table 81. ANOVA on Recognition by Type of Linguistic Item and Time

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Item</td>
<td>1</td>
<td>1.234</td>
<td>1.234</td>
<td>2.121</td>
<td>.150</td>
<td>.029</td>
<td>.301</td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>13.030</td>
<td>6.515</td>
<td>12.722</td>
<td>.000*</td>
<td>.150</td>
<td>.996</td>
</tr>
<tr>
<td>Type of Item * Time</td>
<td>2</td>
<td>.625</td>
<td>.313</td>
<td>.592</td>
<td>.555</td>
<td>.008</td>
<td>.147</td>
</tr>
</tbody>
</table>

*p < .05.

Descriptive statistics (means, SD, and SE for recognition by type of linguistic item and time) are displayed in Table 82. The means show a similar gain for vocabulary and grammar recognition from pretest to immediate posttest.

Table 82. Means, SD, and SE for Recognition by Type of Linguistic Item and Time

<table>
<thead>
<tr>
<th>Type of Linguistic Item</th>
<th>N</th>
<th>Pretest</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M (.SD)</td>
<td>(SE)</td>
<td>(SE)</td>
</tr>
<tr>
<td>Lexical Items</td>
<td>73</td>
<td>.00 (.00)</td>
<td>.007 (.34)</td>
<td>.265 (.085)</td>
</tr>
<tr>
<td>Grammatical Items</td>
<td>73</td>
<td>.05 (.081)</td>
<td>.007 (.39)</td>
<td>.997 (.085)</td>
</tr>
</tbody>
</table>

Note. Maximum score = 1.
To investigate the effect of time on recognition within each type of linguistic item, several One-Way ANOVAs were conducted on word meaning and grammar recognition data with Time as a within-subject factor. Results for lexical items indicated a significant difference between pretest and immediate posttest ($F(1, 72) = 122.325, p = .000$), a significant difference between pretest and delayed posttest ($F(1, 72) = 134.972, p = .000$), and a significant difference between immediate posttest and delayed posttest ($F(1, 72) = 3.972, p = .050$). These results are displayed in Tables 83-85.

Table 83. One-Way ANOVA on Recognition of Lexical Items by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>4.315</td>
<td>4.315</td>
<td>122.325</td>
<td>.000*</td>
<td>.629</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p < .05.

Table 84. One-Way ANOVA on Recognition of Lexical Items by Time: Pretest versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>3.437</td>
<td>3.437</td>
<td>134.972</td>
<td>.000*</td>
<td>.652</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p < .05.

Table 85. One-Way ANOVA on Recognition of Lexical Items by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>.050</td>
<td>.050</td>
<td>3.972</td>
<td>.050*</td>
<td>.052</td>
<td>.503</td>
</tr>
</tbody>
</table>

*p < .05.

Results for recognition of grammatical items indicated a significant difference between pretest and immediate posttest ($F(1, 72) = 8.807, p = .004$), a significant difference between pretest and delayed posttest ($F(1, 72) = 7.290, p = .009$), and no
significant difference between immediate posttest and delayed posttest \((F(1, 72) = .385, p = .537)\). These results are displayed in Tables 86-88.

Table 86. One-Way ANOVA on Recognition of Grammatical Items by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>4.213</td>
<td>4.213</td>
<td>8.807</td>
<td>.004*</td>
<td>.109</td>
<td>.833</td>
</tr>
</tbody>
</table>

*p < .05.

Table 87. One-Way ANOVA on Recognition of Grammatical Items by Time: Pretest versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>7.895</td>
<td>7.895</td>
<td>7.290</td>
<td>.009*</td>
<td>.092</td>
<td>.759</td>
</tr>
</tbody>
</table>

*p < .05.

Table 88. One-Way ANOVA on Recognition of Grammatical Items by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>.573</td>
<td>.573</td>
<td>.385</td>
<td>.537</td>
<td>.005</td>
<td>.094</td>
</tr>
</tbody>
</table>

*p < .05.

In sum, these results show that there was a significant gain from pretest to immediate posttest for both lexical and grammatical items. In addition, this gain was retained over a period of a week since performance was significantly higher on the delayed posttest than on the pretest for both vocabulary and grammar, but there was a significant loss from immediate to one-week delayed posttest in meaning recognition of lexical items. Figure 12 illustrates the pattern for recognition by type of linguistic item.
Figure 12. Recognition by Type of Linguistic Item (Percentage)

Effect of type of linguistic item on written production

The results of the ANOVA for production indicated no significant main effect for Type of Linguistic Item ($F (1, 72) = 1.598, p = .210$), a significant main effect for Time ($F (2, 144) = 34.139, p = .000$), and no significant interaction Type of Linguistic Item x Time ($F (2, 144) = .729, p = .484$). Results for production are displayed in Table 89.

Table 89. ANOVA on Production by Type of Linguistic Item and Time

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Item</td>
<td>2</td>
<td>.091</td>
<td>.091</td>
<td>1.598</td>
<td>.210</td>
<td>.022</td>
<td>.239</td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>2.061</td>
<td>1.030</td>
<td>34.139</td>
<td>.000*</td>
<td>.322</td>
<td>1.000</td>
</tr>
<tr>
<td>Type of Item x Time</td>
<td>2</td>
<td>.039</td>
<td>.019</td>
<td>.729</td>
<td>.484</td>
<td>.010</td>
<td>.172</td>
</tr>
</tbody>
</table>

*p < .05.
Descriptive statistics (means, SD, and SE for production by type of linguistic item and time) are displayed in Table 90. The means indicate a very low gain for both word meaning and grammar production, which was retained over a period of a week.

Table 90. Means, SD, and SE for Production by Type of Linguistic Item and Time

<table>
<thead>
<tr>
<th>Type of Linguistic Item</th>
<th>Pretest</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Lexical Items</td>
<td>73</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Grammatical Items</td>
<td>73</td>
<td>.01</td>
<td>.026</td>
</tr>
</tbody>
</table>

Note. Maximum score = 1.

To investigate the effect of time on production within each type of linguistic item, several One-Way ANOVAs were conducted on word meaning and grammar production data with Time as a within-subject factor. Results for lexical items indicated a significant difference between pretest and immediate posttest ($F(1, 72) = 48.678, p = .000$), a significant difference between pretest and delayed posttest ($F(1, 72) = 78.895, p = .000$), and no significant difference between immediate posttest and delayed posttest ($F(1, 72) = 2.081, p = .153$). These results are displayed in Tables 91-93.

Table 91. One-Way ANOVA on Meaning Production of Lexical Items by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>.543</td>
<td>.543</td>
<td>48.678</td>
<td>.000*</td>
<td>.403</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p < .05.
Table 92. One-Way ANOVA on Meaning Production of Lexical Items by Time: Pretest versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>.685</td>
<td>.685</td>
<td>78.895</td>
<td>.000*</td>
<td>.523</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p < .05.

Table 93. One-Way ANOVA on Meaning Production of Lexical Items by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>.008</td>
<td>.008</td>
<td>2.081</td>
<td>.153</td>
<td>.028</td>
<td>.296</td>
</tr>
</tbody>
</table>

*p < .05.

Results for production of grammatical items indicated a significant difference between pretest and immediate posttest ($F (1, 72) = 16.502, p = .000$), a significant difference between pretest and delayed posttest ($F (1, 72) = 17.812, p = .000$), and no significant difference between immediate posttest and delayed posttest ($F (1, 72) = 2.207, p = .142$). These results are displayed in Tables 94-96.

Table 94. One-Way ANOVA on Production of Grammatical Items by Time: Pretest versus Immediate Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>.631</td>
<td>.631</td>
<td>16.052</td>
<td>.000*</td>
<td>.182</td>
<td>.977</td>
</tr>
</tbody>
</table>

*p < .05.

Table 95. One-Way ANOVA on Production of Grammatical Items by Time: Pretest versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>1.193</td>
<td>1.193</td>
<td>17.812</td>
<td>.000*</td>
<td>.198</td>
<td>.986</td>
</tr>
</tbody>
</table>

*p < .05.
Table 96. One-Way ANOVA on Production of Grammatical Items by Time: Immediate versus Delayed Posttest.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>Eta2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>.089</td>
<td>.089</td>
<td>2.207</td>
<td>.142</td>
<td>.030</td>
<td>.311</td>
</tr>
</tbody>
</table>

*p < .05.

These results show that there was a significant gain from pretest to immediate posttest for lexical and grammatical items, and this gain was retained over a period of a week. Figure 13 illustrates results for production by type of linguistic item.

Figure 13. Production by Type of Linguistic Item (Percentage)

Type of Linguistic Item
- X Grammatical Items
- ◊ Lexical Items
CHAPTER 4: DISCUSSION AND CONCLUSIONS

Discussion

This discussion has been divided into the same sections as the previous chapter addressing the results. The first section discusses the results related to the relationship between experiences of recognition memory and L2 development (i.e., Research Question 1). This will be followed in the next section by discussion of the relationship between levels of awareness and memory experiences, and the relationship between levels of awareness and L2 development (i.e., Research Question 2 and 3). The third section addresses the effects of type of glossing on memory experiences, levels of awareness, L2 development, and text comprehension (i.e., Research Questions 4, 5, 6, and 7). Finally, the last section discusses the effects of type of linguistic item on memory experiences, levels of awareness, and L2 development (i.e., Research Questions 8, 9, and 10). For the reader’s convenience, each research question is repeated before its corresponding discussion.

Experiences of Recognition Memory

The relationship between memory experiences and L2 development

RQ#(1) Is there a relationship between reported experiences of recognition memory and:

(a) learners’ recognition of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

(b) learners’ written production of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

228
If so, will the relationship be maintained over a period of one week?

Research Question 1 received a positive answer with respect to both lexical and grammatical items. The results indicated that remembering positively correlated with word meaning production and recognition at immediate and one-week-delayed posttests, and grammar production and recognition at immediate and one-week-delayed posttests. In addition, there was a negative correlation between guessing and grammar production and recognition at immediate posttest. In other words, the experience of remembering appears to be related to high performance on production and recognition of both lexical and grammatical items, while the experience of knowing is not related to performance on production or recognition tests. Finally, the experience of guessing appears to be related to low performance on immediate grammar production and recognition.

The finding of a positive correlation between remembering and L2 development extend previous findings in the cognitive psychology field (e.g., Conway et al., 1997) to new areas of learning such as second language learning. Moreover, it provides empirical evidence for theoretical models that assign an important role to episodic memory in learning (cf. Tulving, 1993, 1995; Yonelinas, 2002). According to Tulving (1995) both the episodic and semantic memory systems (i.e., the declarative memory system) support learning of new information. Recall performance relies more on the episodic memory system, while recognition relies on both systems. However, the episodic memory system is predicted to facilitate recognition to a greater extent than semantic memory. Importantly, Tulving’s model assumes that the episodic memory system is always
activated in combination with the semantic memory system, while the semantic memory system may be activated without the episodic memory system. Thus, as Craik (2002) has suggested, it may be this encoding in two systems (i.e., rich encoding) versus encoding in one system (i.e., poor encoding) what leads episodic memory to play an important role in learning.

In the SLA field, most researchers agree on the importance of declarative memory in both L2 vocabulary and L2 grammar learning (Ullman, 2001; N. Ellis, 2001, 2005, 2007; Paradis, 1994). However, the issues of whether episodic and semantic memory might play different roles in L2 development, and whether their roles would be the same in vocabulary and grammar development, have not received enough attention in the SLA literature. N. Ellis (2001, 2005, 2007) has argued that episodic memory is crucial in both vocabulary and grammar learning. Studies in cognitive psychology have shown that episodic memory may impact recognition of L1 lexical items (e.g., Gardiner & Java, 1990; Rajaram et al., 2002; Karayianni & Gardiner, 2003), and thus the relationship between remembering and L2 vocabulary would be expected. However, following up on Craik (2002), it may be questioned why memory of specific contextual details should be associated with higher recognition of grammatical structures, which are considered to be characterized by a higher abstraction than lexical items (cf. DeKeyser, 2005). Moreover, the fact that vocabulary and grammar may lead to dissociations in memory in the first language (cf., Ullman, 2001), and in the second language (Peckham, 2000), can make us question whether the relationship between declarative memory and L2 learning may be
different depending on type of linguistic item (e.g., whether episodic memory may be related to lexical learning and semantic memory to grammatical learning).

Findings of the present study bring evidence in support of N. Ellis’s (2001, 2005) view, and confirm that (a) episodic memory as measured by remember responses is positively related to L2 development (i.e., both recognition and production immediately after exposure and one week later) while semantic memory as measured by know responses is not significantly related to L2 development, and (b) the positive relationship between episodic memory and L2 development is significant for both lexical and grammatical items. Thus, the present study has contributed to SLA research on memory and learning by providing empirical evidence for the role of long-term episodic memory in L2 vocabulary and grammar learning.

Levels of Awareness

The relationship between levels of awareness and memory experiences

RQ#(2) Is there a relationship between the level of awareness reported during a reading comprehension task and the experience of recognition memory reported by learners for (i) unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

Research Question 2 received a negative answer with respect to lexical items, and a positive answer with respect to grammatical items. The results indicate that reported noticing of lexical items does not significantly correlate with remembering, knowing, or
guessing experiences. Similarly, the lack of verbal report for lexical items does not appear to correlate with any memory experience. With respect to grammar, the results indicate that understanding, as measured by think-aloud protocols, has a significant positive correlation with remembering and a significant negative correlation with guessing. In other words, the more understanding reported by participants, the more remembering and the fewer guesses. The correlation analysis shows that reported noticing of grammatical items as measured by an online procedure positively correlates with remembering too, although to a lesser extent than understanding.

Previous studies in the cognitive psychology field have supported a significant effect of depth of processing on remembering of L1 lexical items presented in the visual mode (e.g., Gardiner, 1988; Rajaram, 1993; Java, 1994; Gardiner et al., 1994; Richardson-Klavehn & Gardiner, 1995; Gardiner, Java & Richardson-Klavehn, 1996; Yonelinas, 2001; Richardson-Klavehn et al., 2002). Based on findings of cognitive psychology studies, it was expected to find a positive correlation between reported awareness and remembering of L2 lexical items. Therefore, the lack of correlation found in the present study was unexpected. Two factors may account for this finding. First, only evidence for awareness at a low level was found in the present study for vocabulary words. Since depth of processing seems to have a significant impact on remembering, it is plausible that deeper processing, that is, awareness at a level higher than noticing, would have been positively correlated with remembering. Second, a qualitative difference between processing familiar words (i.e., L1 words) versus unfamiliar words (i.e., unfamiliar L2 words) may contribute to explain findings of this study. Previous cognitive
psychology studies (Gardiner & Java, 1990; Rajaram et al., 2002; Karayianni & Gardiner, 2003) have shown that L1 words (i.e., familiar words) lead to significantly more remember responses than non-words (i.e., words that participants have never been exposed to prior to the experiment). Furthermore, Martínez-Fernández (2009a) examined the memory experiences triggered by targeted lexical items (i.e., unfamiliar words that were deleted from the text and glossed and bolded in the margin) and non-targeted lexical items (i.e., familiar words that were present in the text but were not enhanced in any way). Findings of that study showed that participants produced significantly more remember responses for non-targeted items (i.e., familiar and non-enhanced) than for targeted items (i.e., unfamiliar and enhanced) on immediate and one-week-delayed posttests, regardless of whether they were presented under think-aloud or silent conditions. In sum, SLA and cognitive psychology studies have found evidence to support the claim that words whose meaning is unfamiliar to learners are harder to remember, and that prior knowledge may influence how learners process the input to a greater extent than input enhancement techniques. These findings of previous literature could explain why awareness at a low level may be unrelated to episodic memory of unfamiliar L2 words as measured by remember responses.

In addition, awareness at a low level also appears to be unrelated to semantic memory as measured by know responses. The lack of correlation between noticing and both remembering and knowing does not support the idea that noticing (as measured by think-aloud protocols in the present study) involves the conscious registration of the occurrence of a stimulus and its subsequent storage in long-term memory (Schmidt,
However, findings of the present study are consistent with the view that noticing requires rehearsal in working memory prior to encoding in long-term memory. As Robinson (1995b) indicates, “it is possible to briefly notice and permanently or temporarily forget, and to notice and remember over time. More permanent encoding in long-term memory is a consequence of the level of activation of information, itself the result of rehearsal and elaboration” (p. 298). Thus, further research is needed to establish whether higher levels of awareness may be related to long-term episodic and/or semantic memory of unfamiliar L2 lexical items.

Because cognitive psychology studies have mainly examined experiences of recognition memory of L1 lexical items, the literature did not support any prediction with respect to memory experiences of items with grammatical meaning. However, researchers within the attentional framework in SLA seem to assume that awareness plays an important role in episodic memory for both vocabulary words and grammatical structures (e.g., Schmidt, 1994a; Robinson, 1995b). The present study corroborated this view, as findings indicated a positive correlation of understanding and noticing with remembering, and this correlation was higher for understanding (i.e., awareness at a high level) than for noticing (i.e., awareness at a low level). This result also confirms findings of cognitive psychology studies on the effect of levels of processing for L1 lexical items, and provides support for a potential relationship between awareness at a high level and remembering of L2 lexical items.
The relationship between levels of reported awareness and L2 development

RQ#(3) Is there a relationship between the level of awareness reported during a reading comprehension task and:

(a) learners’ recognition of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

(b) learners’ written production of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

If so, is this correlation maintained over a period of one week?

The quantitative analyses on reported levels of awareness provided empirical evidence to give a positive answer to Research Question 3. In the present study, one level of reported awareness of lexical items (i.e., noticing) and two levels of reported awareness of the grammatical structure (i.e., understanding and noticing) were identified. Results indicated that noticing of lexical items positively correlated with immediate and one-week delayed word meaning recognition and production. In addition, the lack of verbal report appeared to be significantly correlated with low performance on immediate word meaning production, although this correlation disappeared on the one-week-delayed posttest, and it was not found on recognition posttests.

The results regarding the grammatical structure showed a positive correlation between understanding and immediate grammar recognition and production of both old and new items. Moreover, this correlation was maintained over a period of a week. Noticing also appeared to positively correlate with immediate and one-week-delayed
grammar recognition and production, and the correlation was a bit lower than for understanding. Finally, the lack of verbal reports had a significant negative correlation with grammar recognition and production, which was higher on delayed than immediate posttests. In sum, these results indicate that participants who reached higher levels of reported awareness of the grammatical structure performed higher on the grammar posttests, and that participants who produced fewer verbal reports tended to perform lower on the grammar posttests.

These findings provide empirical evidence for Schmidt’s noticing hypothesis, and support previous findings on the relationship between levels of awareness and L2 grammar development (Camblor, 2006; Leow, 1997a; Rosa & O’Neill, 1999; Rosa & Leow, 2004), and between noticing and L2 vocabulary development (Guidi, 2009). Leow (1997a) and Rosa and O’Neill (1999) found a significant effect of levels of awareness on intake as measured by performance on immediate posttests. However, these studies did not examine the effect of levels of awareness on performance over time. Rosa and Leow (2004) found support for the differential effect of levels of awareness on both immediate and three-week-delayed posttests. Findings of that study revealed that participants who reached the highest level of awareness (i.e., understanding) experienced a significant loss from immediate posttest to delayed posttest, although they still maintained a significant score gain between pretest and delayed posttest. In the same line, Camblor (2006) found that participants who reached the highest levels of reported awareness significantly outperformed those who did not report any level of awareness on a three-week-delayed posttest. Consistent with these results, the present study indicated that the positive
correlation found between reported levels of awareness and performance on immediate grammar posttests was maintained at the delayed posttest, although the delayed posttest was administered only one week after the treatment, and not three weeks after as in previous studies. Similarly, the positive correlation found between reported noticing and performance on immediate vocabulary tests was maintained one week after the treatment. This result corroborates previous findings reported by Guidi (2009), who found a positive correlation between reported noticing and performance on immediate and three-week-delayed vocabulary posttests. In sum, findings of the present study add to the previous literature to support the claim that awareness is positively related to L2 grammar and vocabulary development.

Type of Glossing

Effect of type of glossing on memory experiences

RQ#(4) Does type of glossing in a reading comprehension task have a differential effect on experiences of recognition memory reported by learners for (i) unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

The answer to Research Question 4 was negative with respect to lexical and grammatical items. The results indicated that condition of exposure did not have a significant impact on the amount of remember, know and guess experiences reported for target lexical items, lexical items overall (i.e., alternate items in the fill-in task plus target items), and target grammatical items.
The fact that the experimental groups did not show higher remembering of target lexical items than the control group contrasts with the finding of higher noticing in the experimental groups than in the control group (cf. discussion of RQ#5 below). In other words, the control group reported less noticing than the glossing conditions but not less remembering. This result, however, is not unexpected taking into account the lack of correlation between reported noticing of lexical items and memory experiences.\(^{29}\) Moreover, a qualitative analysis of the think-aloud protocols revealed that a plausible explanation why remembering was higher than noticing in the control group when compared to other conditions is that participants in the control group, on the one hand, often skipped the unknown words while reading (thus, reported less noticing, from a quantitative perspective), but, on the other hand, engaged in deeper processing (e.g., tried to infer the word meaning, repeated the word form several times) in the few cases when they noticed unknown words. Because participants in the control group were not provided with the meaning of the unfamiliar words, they tended to notice the word form to a greater extent than participants in the glossing conditions, who usually focused on the meaning of the words to understand the text or complete the fill-in task but not so much on the word form. Since participants were asked to recognize word forms and not word meanings in the remember-know task, it is not surprising that participants in the control group were able to remember many of the items they had previously noticed. On an anecdotal note, the analysis of answers given to questions addressing local

\(^{29}\) A correlation analysis was conducted on awareness and memory experiences data excluding data from the control group to find out whether the lack of correlation was due to the relationship between reported noticing and memory experiences in this group. However, that analysis did not yield any significant correlation.
comprehension in the reading questionnaire revealed that some participants from the control group (and only from the control group) provided the targeted word forms in their answers, as they remembered exactly the word forms and recognized the context but did not know the meaning. The following example illustrates the answer given by a participant in the control group to a question addressing local comprehension; the question asked participants to complete the sentence in English based on information from the reading passage, and the target item addressed was arce, ‘maple tree’: “The master was standing outside the house, next to a… arce.”

This finding also shows that remembering may be more related to depth of processing than to external enhancement of the linguistic items since no techniques of input enhancement were used in the control condition. In addition, it brings evidence in support of different levels of processing at one level of awareness (cf. Leow et al., 2008, Leow et al., forthcoming for further discussion). In the present study there was evidence for only one level of awareness of vocabulary words, that is, noticing or awareness at a low level, but participants seemed to process word form and word meaning with different levels of depth.

Findings regarding the effect of type of glossing should be discussed in light of previous findings of the effect of type of glossing on memory experiences reported for vocabulary words. Martínez-Fernández (2009b) investigated the effect of task complexity in lexical tasks during reading under think-aloud and silent conditions. Task complexity was defined by the reasoning demands required by the tasks. In the experimental conditions, participants read a text containing eight blanks, and in order to fill in each
blank they had to choose and write one of two unfamiliar words, which were bolded and
glossed at the margin. In the high reasoning conditions, both options given for each blank
were plausible but only one was correct based on specific contextual clues. In the low
reasoning conditions, one of the two options given was clearly incorrect because it had
the opposite lexical feature (i.e., plus versus minus concrete) to that selected by the
previous word. The results of the study showed that the high reasoning groups
remembered significantly more words overall (i.e., target plus alternate words) than the
control group, while the minus reasoning group was not significantly different from the
high reasoning and control conditions.

Findings of Martínez-Fernández (2009b) suggest that high reasoning conditions
may have a positive effect on remembering when compared to a control condition. This
finding was not corroborated by the present study, where the minus incidental condition
(i.e., a high reasoning condition) did not remember significantly more words than the
control group. These conflicting findings, however, may be explained by the presence of
a production component in the fill-in task used by Martínez-Fernández (2009b), which
might have an effect on processing of word forms and episodic memory. Moreover, while
the experimental conditions were not significantly different in Martínez-Fernández’
(2009b) study, the experimental conditions were significantly different in the present
study. In other words, the distance between high and low reasoning conditions turned out
not to be significant, while the distance between minus and plus incidental conditions
(i.e., high reasoning versus no reasoning conditions) appeared to be significant when
examining memory experiences of lexical items overall. However, in contrast with the
minus incidental group, the plus incidental group was not exposed to alternate words in the present study. Therefore, no conclusions can be made regarding the effect of reasoning and evaluation triggered by minus incidental conditions on remembering. In order to investigate this issue, future studies will need to employ research designs where task variables that are not the focus of the research (i.e., production of items, exposure to alternate items, etc.) are kept constant across tasks.

To summarize up to this point, condition of exposure did not appear to impact experiences of recognition memory of target lexical items, although a significant effect was found when taking into account lexical items overall. Whether this effect was a result of a quantitative aspect of the task (i.e., exposure to a higher amount of items) or a consequence of a qualitative aspect (i.e., evaluation or reasoning triggered by the minus incidental condition) needs to be further investigated. Finally, with respect to memory experiences of grammatical items, no significant effect was found for condition of exposure. This result is not surprising, as type of glossing did not have an impact on reported levels of awareness (cf. discussion on RQ #5 below), and reported levels of awareness were found to correlate with memory experiences. In conclusion, future research needs to investigate a variety of types of tasks, and establish whether type of task may have a differential effect on experiences of recognition memory of lexical and grammatical items. Findings of the present study supported the view that internal processes (e.g., experiences of recognition memory) may occur independently of external conditions (cf. Leow, 2001b; Leow et al, 2003), and that saliency may be internally generated (cf. Sharwood Smith, 1991).
Effect of type of glossing on levels of reported awareness

RQ#(5) Does type of glossing in a reading comprehension task have a differential effect on levels of awareness reported by learners for (i) unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

The answer to Research Question 5 is partially positive with respect to lexical items, and negative with respect to grammatical items. The results indicated that type of glossing had a significant effect on reported noticing of lexical items. The post-hoc Scheffé tests revealed that the amount of noticing of target lexical items was significantly higher in the experimental groups (i.e., minus and plus incidental groups) than in the control group, while the amount of no verbal report was significantly higher in the control group. The experimental groups were not significantly different in amount of noticing of target lexical items. However, when taking into account lexical items overall (i.e., alternate items in the fill-in task plus target items), the analyses indicated that the amount of noticing of lexical items was significantly higher in the minus incidental condition than in the plus incidental and control conditions, and that reported noticing of lexical items was significantly higher in the plus incidental condition than in the control group.

With respect to the grammatical structure, the analysis of reported awareness as measured by think-aloud protocols indicated no significant difference between conditions of exposure in amount of reported noticing or no verbal report. However, the amount of reported understanding was significantly higher in the minus incidental group than in the
control group, while there was no significant difference between the minus and plus incidental groups, or between the plus incidental and control groups. When taking into account levels of awareness reported by all participants (i.e., participants in both think-aloud and silent conditions) through online and/or offline procedures, the results indicated no significant difference between groups in amount of understanding, noticing or no verbal report. Overall, very few participants were able to verbalize the grammatical rule underlying the target structure, regardless of the type of measure used. For this reason, when analyzing data from the whole sample, the amount of reported understanding in the experimental conditions turned out not to differ statistically from that in the control group (where cases of reported understanding were not found). Importantly, cases of reported understanding were found in both the minus and plus incidental conditions. This suggests that the fill-in task, which was predicted to direct learners’ attention to the contrast between past and present subjunctive forms to a greater extent, did not have any effect when compared to simple exposure to glosses of the past and present subjunctive forms.

In sum, while type of glossing (i.e., glosses embedded in a fill-in task versus traditional glosses versus no gloss) appeared to have a positive effect on noticing of lexical items, the present study did not provide enough evidence to support a positive effect on levels of awareness of a grammatical structure. These findings should be discussed in the context of previous studies on input enhancement that have attempted to empirically investigate noticing (Leow, 2001b; Leow et al. 2003; Bowles, 2003; Bowles, 2004; Guidi, 2009; Martínez-Fernández, 2008a). Results of the present study corroborate
findings of previous studies examining the effect of input enhancement techniques on reported noticing of grammatical items within the textual enhancement strand (Leow, 2001b; Leow et al. 2003; Bowles, 2003). Leow (2001b) included think-aloud protocols to investigate reported noticing of linguistic forms under enhanced versus non-enhanced conditions, and found no significant difference between groups, a finding that was further confirmed by Leow et al. (2003) and Bowles (2003). Thus, these studies do not support the claim that input enhancement techniques are effective to direct learners’ attention to and facilitate noticing of grammatical forms. Similarly, in the present study the amount of reported noticing of grammatical items did not differ statistically by type of glossing, regardless of the procedure used to measure noticing. Although the minus incidental group seemed to lead to significantly more reported understanding (as measured by think-aloud protocols) than the control group, cases of understanding were almost anecdotal and the difference turned out not to be significant when taking into account data from a larger sample of participants. Moreover, the qualitative analyses of the think-aloud protocols suggest that understanding and noticing of grammatical items may occur in any condition of exposure, depending on how learners interact with the material.

Within the glossing strand, the present study partially contradicts findings reported by Guidi (2009), who investigated reported noticing of lexical and grammatical items as measured by think-aloud protocols under gloss and no-gloss conditions. Results of that study indicated no effect of glossing on reported noticing. Although the present study confirms this finding with respect to grammatical items, it does not corroborate it with respect to lexical items. Results of this study suggest that glossing has a significant
positive effect on reported noticing of lexical items when compared to a control condition, thus confirming previous glossing studies on noticing of vocabulary words (Bowles, 2004). This finding suggests that glossing might be more effective to direct learners’ attention to lexical items than to grammatical items in the context of reading a text. However, to date very few studies on glossing have attempted to measure noticing, and further research would be necessary in order to confirm the effectiveness or lack of effectiveness of glossing on noticing of lexical and grammatical items. Beyond the differences between glossing studies, findings of Guidi (2009), Bowles (2004) and the present study reveal that noticing occur in both gloss and no-gloss conditions of exposure. Thus, all studies provide evidence to support the claim that noticing depends on how the learner interacts with the material, regardless of whether or not condition of exposure may also have an impact on noticing (cf. also Leow, 2001; Leow et al, 2003).

Finally, in line with previous findings reported by Guidi (2009) and Bowles (2004), there was almost no evidence of awareness at a higher level than noticing for lexical items in the present study. Overall, participants in the experimental conditions focused on the meaning of the lexical items to gain deeper understanding of the text and/or choose the words that made more sense within the context to complete the fill-in-the-blank task. They did not seem to focus on the lexical items as language per se, or make explicit connections between word forms and meanings. This finding contrasts with Martínez-Fernández’ (2008a) findings. In that study, exposure to traditional glosses versus exposure to glosses embedded in a fill-in task had a differential impact on types of noticing as measured by think-aloud protocols. Specifically, exposure to traditional
glosses led to reported noticing of meaning (i.e., noticing of one word aspect) while exposure to glosses embedded in a fill-in task led to a significantly higher amount of reported noticing of both word form and meaning (i.e., noticing of two word aspects). In addition, although the traditional gloss and fill-in conditions were not significantly different in meaning recognition and production of target words, the fill-in condition was the only one to significantly outperform the control group. Based on previous studies supporting a relationship between awareness and learning, Martínez-Fernández (2008a) suggested that reported noticing of two word aspects might involve awareness at a higher level than reported noticing of one word aspect, a suggestion that needed to be confirmed by further research. In the present study, however, the qualitative analysis of the think-aloud protocols revealed that both conditions of exposure led to reported noticing of meaning, and, to a lesser extent, to reported noticing of both word form and meaning. In explaining these conflicting results, it is important to take into account that the fill-in task used by Martínez-Fernández (2008a) replicated the one employed in Hulstijn and Laufer’s (2001) experiments, which involved production of the target words, that is, participants had to write the target words to fill in the blanks, instead of underlining the correct options like in the present study. Thus, it seems plausible that this production component might have led learners to isolate the target words, notice the word forms to a greater extent, and make stronger connections between form and meaning. The present study, however, did not find evidence for different levels of awareness of lexical items, and did not support a differential impact of traditional glosses versus glosses embedded in a fill-in task on reported awareness.
Effect of type of glossing on L2 development

RQ#(6) Does type of glossing in a reading comprehension task have a differential effect on:

(a) learners’ recognition of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

(b) learners’ written production of (i) the meaning of unfamiliar Spanish lexical items, and (ii) an unfamiliar Spanish grammatical structure?

If so, will the effect be maintained over a period of one week?

Research Question 6 received a partially positive answer with respect to lexical items, and a negative answer with respect to grammatical items. The results regarding the effect of type of glossing on meaning recognition and production of target lexical items indicated that the experimental groups (i.e., minus and plus incidental) significantly outperformed the control group (i.e., the most incidental) in both measures, and this effect was maintained over a period of a week. The experimental groups, however, were not significantly different in their ability to translate and recognize target words. When examining the effect of type of glossing on meaning production and recognition of lexical items overall (i.e., alternate words in the fill-in task plus target words), the results revealed that the minus incidental group (which had been exposed to both target and alternate words) significantly outperformed both the plus incidental and control groups (which had been exposed only to target words) on meaning recognition posttests.
However, the experimental groups were not significantly different in performance on meaning production posttests.

The results regarding the effect of type of glossing on recognition and production of the target grammatical structure showed that the experimental groups significantly improved from pretest to immediate posttest, and retained this gain over a period of a week, while the control group did not significantly change over time. However, the experimental groups did not significantly outperform the control group in grammar recognition, and only differed from the control group in the gain of old items from the production pretest to the immediate production posttest. Because overall performance on recognition was higher than on production, the difference between the experimental conditions and the control group revealed by the gain in production may have been smaller in grammar recognition. In addition, the gain in grammar production was only experienced for old items, which suggests that types of glossing were not effective to facilitate deep learning of a grammatical structure, which would involve generalization of the grammatical rule to new items. Overall, the main effect for group was not significant on either recognition or production. Thus, it cannot be concluded that conditions of exposure had a significant impact on L2 grammar development.

The findings of the effects of type of glossing on L2 development clearly reflect the positive effect found for type of glossing on noticing of lexical items, and the lack of effect on noticing of grammatical items. Moreover, findings regarding effects on lexical development support previous studies within the glossing strand (e.g., Jacobs, Dufon, & Hong, 1994; Bowles, 2004; Hulstijn, 1992; Hulstijn et al., 1996; Watanabe, 1997).
Overall, these studies show that glossing facilitates vocabulary learning although the gain is generally low. An interesting finding of the present study is the lack of significant difference between the minus and plus incidental conditions in intake of target lexical items. This result contradicts findings of previous studies supporting the Involvement Load Hypothesis (Hulstijn & Laufer, 2001; Keating, 2008; Kim, 2008). These studies have supported a positive effect of glosses embedded in a fill-in task (i.e., minus incidental condition), which are predicted to induce higher involvement on the learners’ part, over traditional glosses (i.e., plus incidental condition). However, although the fill-in and gloss conditions were assumed to differ only in the evaluation processes induced by the former condition and not by the latter, previous studies have employed a fill-in task that involved writing the target items, that is, a production component that could be conflated with the evaluation component, and have a strong impact on lexical development. Moreover, researchers did not employ process measures to ensure that conditions led to the involvement predicted.

The lack of difference found between fill-in and gloss conditions in the present study supports previous findings reported by Martínez-Fernández (2008a). However, some methodological differences make difficult to compare findings of that study with previous studies. Although Martínez-Fernández’ (2008a) study also used a fill-in task that involved writing, the target items occurred four times within the experimental text instead of once as in previous studies. Thus, amount of exposure might have interacted with quality of exposure, and minimized a potential difference between conditions. In conclusion, further research was needed to confirm whether fill-in conditions without
involving production are more conducive to vocabulary development than gloss conditions when participants are exposed to the target words only once. Furthermore, none of the previous studies comparing fill-in and gloss conditions (Hulstijn & Laufer, 2001; Keating, 2008; Kim, 2008; Martínez-Fernández, 2008a) have controlled for the distance between the target and alternate options provided to fill in the blanks, and performance on alternate items was not measured.

The present study addressed these methodological issues, and found that both experimental conditions significantly outperformed the control group, while no significant difference was found between the experimental conditions in immediate recognition and production of target lexical items. Interestingly, when examining lexical development overall (i.e., performance on target plus alternate items), the results show that the fill-in condition led to significantly higher meaning recognition than the gloss condition. However, it should be noted that this superiority of the fill-in condition might be explained by the mere exposure to a higher number of lexical items rather than to the evaluation processes induced by the fill-in task. In other words, participants might learn more lexical items but not necessarily learn lexical items better. In this sense, the effect of quantity of exposure over quality of exposure might also explain why the fill-in condition outperformed the gloss condition in meaning recognition but not in meaning production, since production tests measure deeper and qualitatively higher development than recognition tests. In addition, the fill-in condition spent significantly more time in reading the texts than the gloss condition, which reduces the advantage of fill-in conditions over gloss groups in learning lexical items overall.
Finally, the think-aloud protocols indicated that participants in the fill-in condition engaged in evaluation processes while participants in the gloss condition did not. This high involvement on the learners’ part did not appear to affect intake of target items. However, findings on the effect of time suggest that high involvement might have a positive effect on retention of target words when compared to low involvement. The results show that the minus incidental group retained the gain in meaning recognition of target items from immediate to delayed posttest. In contrast, the plus incidental group experienced a significant loss in meaning recognition from immediate to delayed posttest, although it retained a significant gain with respect to performance on the pretest. This result supports findings reported by Rott (2005), Kim (2008), and Martínez-Fernández (2008a), who found lower retention of target items in gloss conditions than in conditions inducing evaluation processes at delayed posttests. The significant loss in the gloss group was not found for meaning production. However, this may be explained by a practice effect on the meaning production delayed posttest, as a result of completing meaning production and recognition tests during the second session of the experiment. The fact that the control group did not improve from pretest to immediate and delayed posttests but showed a significant gain in meaning production from immediate posttest to delayed posttest seems to support this explanation. Since overall performance on recognition was higher than on production, the effect of practice on recognition might have been smaller. Finally, although the fill-in group showed higher retention in meaning recognition than the gloss condition, it should be noted that it spent significantly more time on task, and thus the difference in retention could also be attributed to this factor.
To summarize up to this point, findings of this study led to the following conclusions: (a) glossing had a positive effect on vocabulary development when compared to a control condition, (b) glosses embedded in a fill-in task did not have an impact on intake of target words when compared to traditional glossing, but might have led to a positive effect on retention, (c) glosses embedded in a fill-in task led to recognition of a higher amount of words overall, which might be explained by exposure to additional lexical items, and (d) the fill-in condition spent significantly more time to complete the reading comprehension task, which might have contributed to the beneficial effect on retention of target words, and on intake of words overall.

With respect to the effect of conditions of exposure on grammar recognition and production, the present study corroborates findings of previous studies that do not support a positive effect of textual enhancement techniques on L2 grammatical development (Shook, 1994; Alanen, 1995; White, 1998; Overstreet, 1998; Izumi, 2002; Leow, 1997b, 2001b; Leow et al., 2003; Bowles, 2003). In addition, the finding of no effectiveness of glossing to facilitate grammatical development partially supports findings reported by Guidi (2009), who found no effect of glossing on recognition or production of the Spanish impersonal marker *se* but a positive effect on immediate recognition of the Spanish present perfect. This positive effect was explained by the combination of high frequency (i.e., ten occurrences within the input) and relative simplicity of the grammatical structure. The target grammatical structure of the present study also occurred ten times within the experimental text but was more complex than the present
perfect, which might explain to some extent the lack of effect of glossing, based on Guidi (2009).

In sum, the present study showed that the combination of textual enhancement and provision of a gloss did not significantly facilitate recognition or production of a complex grammatical structure. Rather, as Leow et al. (2003) concluded, reported levels of awareness appeared to be a more crucial factor than the external conditions of exposure to explain L2 grammar learning in the present study. As discussed above (cf. discussion of RQ#3 and RQ#5), those learners who showed understanding of the grammatical structure performed higher on grammar recognition and production posttests, independently of the type of glossing they were exposed to.

Effect of type of glossing on text comprehension

RQ#(7) Does type of glossing in a reading comprehension task have a differential effect on learners’ text comprehension?

Research Question 7 received a partially positive answer with respect to the effect of type of glossing on comprehension of local ideas (i.e., ideas expressed by the target lexical items), and a negative answer with respect to the effect on comprehension of global ideas (i.e., ideas not expressed by the target items). The results indicated that performance on the reading comprehension questionnaire containing target lexical items was significantly higher on global ideas than on local ideas, and the experimental groups significantly outperformed the control group on local comprehension. No significant
difference was found between the experimental groups in local comprehension, and the experimental and control groups did not differ statistically in global comprehension. Finally, the analysis of global comprehension of the text containing target grammatical items indicated no significant difference between conditions of exposure.

Previous studies on glossing have generally found evidence to support the effectiveness of glosses in reading comprehension (Davis, 1989; Jacob, 1994; Hee Ko, 2005; Bell and Le Blanc, 2000; Bowles, 2004; Guidi, 2009). Thus, the finding of a positive effect of glossing on local comprehension in the present study supports previous literature. Additionally, it was important to test whether glosses embedded in a fill-in task might negatively impact text comprehension, have a beneficial effect or have no effect. Among the studies that have measured the effect of fill-in conditions on text comprehension, one study has found that glosses embedded in a fill-in task did not impact text comprehension, as this condition was not significantly different from the gloss condition or the control group, and the gloss condition was the only one to significantly outperform the control group (Martínez-Fernández, 2008a). In contrast, Martínez-Fernández (2009b) found evidence to support the claim that fill-in conditions might positively impact text comprehension when compared to control conditions. These conflicting results may be explained by the different measure used to assess text comprehension (i.e., written recall task versus reading questionnaire with open questions). The present study confirms results of these two previous studies, since findings show that the fill-in condition significantly outperformed the control group on local comprehension, and the difference between fill-in and gloss conditions was not
significant. Moreover, types of glossing did not have a detrimental effect on global comprehension, as glossing conditions were not significantly different from the control group. These findings show that simultaneous attention to form and meaning (i.e., lexical item and text content) as measured by think-aloud protocols in minus and plus incidental conditions did not have a negative effect on text comprehension or L2 learning, as some researchers have suggested (cf. Han et al, 2008).

Overall, global comprehension was neither high nor low (i.e., participants demonstrated comprehension of approximately half of the global ideas included in the reading questionnaires). Although a systematic analysis of reading strategies was not addressed in this study, the think-aloud protocols revealed that participants used a variety of strategies regardless of condition. Many participants seemed to read the text in a bottom-up fashion, reading aloud or translating word by word, which confirms findings reported by Guidi (2009) and Leow et al (2008). However, the use of global reading strategies such as making inferences, activating prior knowledge, making connections between paragraphs or recapitulating ideas was also well documented in all conditions of exposure.

Type of Linguistic item

Effect of type of linguistic item on memory experiences

RQ#(8) Does type of linguistic item (i.e., lexical versus grammatical) in a reading comprehension task have a differential effect on experiences of recognition memory reported by learners?
Research Question 8 received a positive answer. The results indicated that lexical items received significantly more remember and know responses than grammatical items, which corroborate Peckham’s (2000) results regarding the difference between grammar and vocabulary on memory experiences. This finding is particularly interesting since no evidence to support a differential effect of type of linguistic item on reported awareness of target items was found in the present study (cf. discussion of RQ#9 below). Although it is generally assumed that vocabulary and grammar stimulate different cognitive processes, and lexical items are easier to be noticed while processing the input, results of reported awareness did not appear to support this theoretical view. However, results regarding memory experiences bring evidence in support of the idea that lexical items are significantly easier to be recognized than grammatical items. Thus, this finding suggests the possibility that grammatical items were noticed to the same extent as lexical items, but were not subsequently stored or further processed into the language system. As discussed above (cf. discussion of RQ#2) and following Robinson (1995b), noticing requires rehearsal in working memory prior to encoding in long-term memory; from this perspective, what is noticed may be encoded in episodic memory but not everything that is noticed is processed in long-term memory.

In conclusion, although type of linguistic item can certainly have an effect on reported noticing, as previous SLA studies have shown (e.g., Leow et al., 2003; Guidi, 2009), findings of the present study suggest that the crucial difference between processing vocabulary words and grammatical structures may not lie at the stage between attention and noticing but rather at the stage between noticing and further processing into
episodic and semantic memory. However, since no previous empirical studies have addressed the relationship between noticing and episodic and semantic memory of vocabulary words versus grammatical structures, further research is needed in order to consolidate this finding.

Effect of type of linguistic item on reported levels of awareness

RQ#(9) Does type of linguistic item (i.e., lexical versus grammatical) in a reading comprehension task have a differential effect on levels of awareness reported by learners?

Research Question 9 did not receive a positive answer. Results of this study indicated that amount of reported noticing and amount of no verbal report as measured by think-aloud protocols did not differ statistically by type of linguistic item. Moreover, the analyses with reported levels of awareness of the grammatical structure as a factor indicated that the group of participants who reported awareness of the target grammatical structure at the highest level (i.e., the Understanding group) as measured by think-aloud protocols reported significantly more noticing of lexical items than the group of participants who produced a low amount of verbal reports of the target grammatical structure (i.e., the Low Amount of Verbal Report group). Additionally, the Low Amount of Verbal Report group showed an amount of no verbal report of lexical items significantly higher than the Understanding group. The group of participants who reported awareness of the target grammatical structure at a low level (i.e., the Noticing group) was not significantly different from the other groups in amount of reported
noticing and no verbal report of lexical items. In sum, these results indicate that participants who reported awareness of the grammatical structure at a high level also reported awareness of a higher number of lexical items, and participants who produced fewer verbal reports of the grammatical structure reported awareness of fewer lexical items. This result should be taken with caution, however, due to the low number of participants in the Understanding group.

Previous literature in SLA has generally assumed that learners have a priority for processing lexical information over grammatical information when they are exposed to the input (cf. VanPatten, 1994, 1996, 2004). To date, however, there is not enough empirical evidence to support this view. Guidi (2009) examined the effect of vocabulary versus two grammatical structures differing in their degree of inherent difficulty (i.e., Spanish present perfect and impersonal marker ‘se’) on reported noticing. Findings of that study revealed (a) no significant difference between reported noticing of lexical items and reported noticing of the easy grammatical form (i.e., present perfect), (b) an amount of reported noticing of lexical items significantly higher than the amount of reported noticing of the difficult grammatical form (i.e., impersonal marker), and (c) no significant difference between reported noticing of easy versus difficult grammatical forms. Thus, Guidi (2009) provided partial support for the claim of a priority for processing vocabulary words over grammatical information while reading. The lack of difference between lexical items and the present perfect forms in reported noticing was unexpected, as lexical items were considered more salient, less abstract, and less complex than grammatical items. The researcher concluded that a frequency effect (i.e., multiple
exposure to one grammatical form versus single exposure to multiple lexical items) might have interacted with the relative simplicity of the easy grammatical form, and played a role in reported noticing. In other words, high frequency could explain the noticing of the present perfect, while the high number of unknown items presented with low frequency might explain the results for noticing of vocabulary items.

The finding of no difference between reported noticing of lexical and grammatical items in the present study corroborates to some extent findings reported by Guidi (2009). Like in this study, participants were exposed to ten target lexical items and encountered the present subjunctive form ten times. However, the degree of inherent difficulty of the present subjunctive to express future actions is higher than that of the present perfect in terms of complexity of the relationship between meaning and form (i.e., the present subjunctive may encode different meanings, and future actions may be encoded in different forms), and in terms of saliency (i.e., the present subjunctive is considered to be less salient since it consists of only one morpheme - e.g., *coma* - while the present perfect consists of two morphemes - e.g., *ha comido* -). Therefore, results of the present study contradict Guidi’s (2009) finding of the difference between reported noticing of lexical items and noticing of a grammatical form that is inherently more difficult than the present perfect and that occurs ten times within the text. In other words, if frequency played a role in noticing of the grammatical structure in the present study, it would be necessary to explain why it did not play a role in noticing of the difficult grammatical structure in Guidi’s (2009) study. A possible explanation would be that learners in her study tended to focus on the lexical part of the difficult grammatical items since they were unfamiliar
vocabulary words. In the present study, however, the grammatical items were familiar verbs, which might have allowed learners to focus on the grammatical information to a greater extent.

Guidi (2009) also suggested that the present perfect forms might have been more salient than what was predicted in her study, given that the present perfect consists of two morphemes while vocabulary words consist of only one morpheme. In the present study, both lexical and grammatical items consisted of one morpheme. Therefore, it might be the case that vocabulary words and grammatical items turned out to have similar saliency, which could explain why lexical and grammatical items were noticed to a similar extent (cf. Leow et al., 2003 for a discussion of the effect of saliency on noticing).

To conclude, the present study did not provide evidence to support a differential effect of type of linguistic item, that is, vocabulary versus grammar, on reported awareness of target items as measured by an online procedure. Although the level and the amount of reported awareness overall varied across individuals, participants appeared to report similar amount of awareness for grammatical and lexical items, and those who reported higher awareness of the grammatical structure also reported higher awareness of vocabulary words. The lack of effect for type of linguistic item is unexpected since previous SLA studies generally claim that not all language is processed alike, and that there is a priority for processing of lexical information while reading (e.g., VanPatten, 1994, 1996, 2004). The unexpected finding of the present study may be explained as a result of a frequency effect (i.e., high frequency of the grammatical structure versus low frequency of vocabulary words) in combination with high familiarity of the verbs
selected as target grammatical items, and as a result of similar saliency of vocabulary words and grammatical forms. However, another plausible explanation may be that type of linguistic item did not have an effect on noticing but on further processing that might have taken place immediately after noticing. Findings on experiences of recognition memory seem to support this possibility (cf. discussion of RQ #8).

**Effect of type of linguistic item on production and recognition**

RQ#(10) Does type of linguistic item (i.e., lexical versus grammatical) in a reading comprehension task have a differential effect on (a) learners’ recognition, and (b) learners’ written production? If so, will the effect be maintained over a period of one week?

Research Question 10 did not receive a positive answer. The results indicate that there was a significant gain from pretests to posttests for recognition and recognition of both vocabulary and grammar. The gain in recognition, and to a greater extent the gain in production, was very low, and no significant differences were found by type of linguistic item. In other words, the percentage of lexical items that participants were able to translate and recognize was not significantly different from the percentage of grammatical items that participants were able to produce and recognize in the grammar posttests. On the one hand, these results must be taken with caution due to the low power of the statistical analyses. On the other hand, it should be noted that recognition or production of 10% of target lexical items minimally involves intake or learning of one
vocabulary word, while recognition or production of 10% of grammatical items cannot involve any learning of the grammatical structure. Therefore, learning measures are difficult to compare across the two types of linguistic items examined in this study. Yet, it is an interesting finding that participants gave accurate responses to almost the exact same percentage of items in vocabulary and grammar posttests, a result that reflects the pattern found for the effect of type of linguistic item on reported noticing. Moreover, even though the low performance on grammar recognition and production posttests cannot be considered to reflect any learning of the grammatical structure, findings of this study showed a significant gain from grammar pretests to posttests, which may be explained by the relationship found between noticing of target forms and performance on recognition and production posttests (cf. discussion of RQ #3).

Conclusions

The present study sought to expand findings of previous research in cognitive psychology and SLA by investigating experiences of recognition memory (i.e., remembering and knowing) in relation to second language learning, levels of awareness, condition of exposure, and type of linguistic item. This study was motivated by several research gaps in the SLA literature. On the one hand, the role of experiences of recognition memory has not received enough attention in SLA, and the relationship between memory experiences and awareness and learning has not been empirically investigated. On the other hand, SLA studies on input enhancement suggest that different types of exposure to the input may affect learners’ processing and learning but have
generally failed to measure cognitive processes, and findings are not conclusive. Moreover, the differential impact of vocabulary versus grammar on both cognitive processes and L2 learning needs to be further investigated within the input enhancement strand. Findings of this study contributed to shed light on these issues, and led the researcher to draw the conclusions presented below.

In the first place, the empirical evidence provided in the study indicated a significant relationship between experiences of recognition memory and L2 development. Specifically, remembering appears to be positively related to L2 development while knowing is not significantly related to recognition or production of the target items. Furthermore, remembering seems to have a positive correlation not only with vocabulary learning, as could be predicted by cognitive psychology research, but also with grammar learning. Based on cognitive psychology research, reports of remembering reflect episodic memory (Tulving, 1983, 1985). Thus, these results support N. Ellis’s (2001, 2005) view on the important role that episodic memory plays in both L2 vocabulary and L2 grammar learning.

In the second place, findings of the present study shed light on the relationship between awareness, memory experiences and L2 learning. Levels of awareness reported for the target grammatical structure positively correlated with remembering of grammatical items, and with grammar recognition and production at immediate and one-week-delayed posttests. Moreover, the correlations between understanding, on the one hand, and remembering, recognition, and written production, on the other, were higher than the correlations between noticing and remembering, recognition, and written
production. Therefore, these findings support previous cognitive psychology studies examining the effects of levels of processing on memory experiences of L1 lexical items (e.g., Gardiner, 1988; Rajaram, 1993; Java, 1994; Gardiner et al., 1994; Richardson-Klavehn & Gardiner, 1995; Gardiner, Java & Richardson-Klavehn, 1996; Yonelinas, 2001; Richardson-Klavehn et al., 2002), as well as previous SLA literature addressing the role of levels of awareness on L2 grammar development (Camblor, 2006; Leow, 1997a; Rosa & O’Neill, 1999; Rosa & Leow, 2004). In addition, it provides empirical evidence to support the relationship between awareness and long-term episodic memory that SLA researchers have suggested from a theoretical perspective (e.g., Robinson, 1995b).

Findings pertaining to the relationship between awareness, memory experiences and learning of L2 vocabulary words yielded a more complex picture. Noticing of lexical items positively correlated with meaning recognition and production at immediate and one-week-delayed posttests, thus confirming the relationship between noticing and vocabulary development found in previous studies (Guidi, 2009). However, reported noticing did not appear to correlate with any experience of memory. This unexpected finding was explained taking into account two factors: (a) the effect of levels of awareness on memory experiences, and (b) the effect of familiarity versus unfamiliarity on memory experiences. In the present study, only evidence for awareness at a low level was found for vocabulary words. Therefore, based on previous research, it is plausible that awareness at a level higher than noticing would have been positively correlated with remembering. In addition, previous cognitive psychology and SLA studies (Gardiner & Java, 1990; Rajaram et al., 2002; Karayianni & Gardiner, 2003; Martínez-Fernández,
2009a) indicate that familiar words (e.g., L1 words or familiar L2 words) lead to significantly more remembering than unfamiliar words (e.g., non-words or unfamiliar L2 words). The combination of these two factors (i.e., awareness at a low level and unfamiliarity) might explain why noticing did not correlate with remembering of vocabulary words in the present study. With respect to unfamiliar L2 vocabulary words, evidence of awareness at a high level might be necessary in order to find a significant correlation with remember responses.

Importantly, since noticing of lexical items did not correlate with either remembering or knowing, the present study does not support Schmidt’s (1994a) idea that noticing involves the conscious registration of a stimulus and its subsequent storage in long-term memory. Findings of this study are more in line with Robinson’s (1995b) view of noticing where noticing is partially distinguished from long-term memory, and long-term memory is viewed as a consequence of the level of processing.

Third, findings of the effect of type of glossing on reported awareness, memory experiences, L2 development, and text comprehension, led to a deeper understanding of the effectiveness or lack of effectiveness of pedagogical techniques involving glossing. Providing L1 translation glosses or L1 translation glosses embedded in a fill-in task for grammatical items (i.e., reading tasks involving plus versus minus incidental conditions, respectively) did not appear to have a positive effect on reported levels of awareness, memory experiences, intake or retention over a period of one week. The lack of effectiveness of condition of exposure in grammar recognition and production corroborates findings of previous studies that do not support a positive effect of textual
enhancement techniques on L2 grammatical development (Shook, 1994; Alanen, 1995; White, 1998; Overstreet, 1998; Izumi, 2002; Leow, 1997b, 2001b; Leow et al., 2003; Bowles, 2003). Based on a glossing study conducted by Guidi (2009), it could be argued that glossing might interact with frequency of exposure to facilitate learning of easy grammatical structures but not difficult grammatical structures, such as the one selected in the present study. In any case, findings of this study suggest that reported levels of awareness might be a factor more important than the external conditions of exposure to explain L2 grammar learning, as has been suggested by previous research (e.g., Leow et al., 2003). Learners who reported understanding of the grammatical structure performed higher on the grammar recognition and production posttest tests, independently of the type of glossing they were exposed to.

In contrast, providing L1 translation glosses and L1 translation glosses in an embedded fill-in task for lexical items seem to have a positive effect on reported noticing and L2 vocabulary learning, although the vocabulary gain was very low, as has been found in previous glossing studies (Jacobs, Dufon, & Hong, 1994; Bowles, 2004; Hulstijn, 1992; Hulstijn et al., 1996; Watanabe, 1997). Although the experimental conditions showed more reported noticing than the control group, they did not show higher remembering of target lexical items. However, this result was not surprising taking into account the lack of correlation between reported noticing of lexical items and memory experiences, and given the fact that what is noticed is not necessarily remembered.
The effect of traditional glosses was not significantly different from the effect of glosses embedded in a fill-in task on reported noticing, memory experiences, or intake of target lexical items. This finding provides evidence against previous studies supporting the Involvement Load Hypothesis (Hulstijn & Laufer, 2001; Keating, 2008; Kim, 2008), which predicts that tasks inducing evaluation of lexical items (e.g., reading a text with glosses embedded in a fill-in task) will be more beneficial for vocabulary learning than tasks that do not induce evaluation (e.g., reading a text with traditional glosses). A plausible explanation of these conflicting findings may lie in the fact that the fill-in task in previous studies involved writing the lexical items while in the present study it involved underlining the items selected to fill in the blanks. This difference led to conclude that production might have played a role in vocabulary learning. Different from previous studies, the present study examined the effect of type of glossing not only on target lexical items but also on alternate lexical items (i.e., items that were presented only in the minus incidental condition). The results with respect to lexical items overall indicated that the minus incidental group significantly outperformed the plus incidental group in amount of reported noticing, amount of reported remembering, intake, and retention. However, from these findings it cannot be concluded that the positive effect of the minus incidental condition was due to the evaluation processes induced by the fill-in task. On the one hand, time on task may have contributed to this result since the minus incidental condition spent significantly more time in completing the reading task than the plus incidental condition. On the other hand, the plus incidental task did not include exposure to the alternate lexical items. Therefore, the superiority of the minus incidental
condition may be explained by the simple exposure to a higher number of glossed words. The qualitative analysis of the think-aloud protocols suggest that participants in both conditions were processing the vocabulary words for their meaning, and there was no evidence of explicit connections between word meaning and form. This result might explain the lack of difference between conditions in intake of target lexical items. Additionally, the think-aloud protocols revealed that the fill-in condition led to evaluation processes while the traditional glossing condition did not. This higher involvement induced by the fill-in condition might account for the positive effect found for this condition on retention of target items with respect to the traditional glossing condition, which experienced a significant loss in meaning recognition from immediate to delayed posttest (cf. Martínez-Fernández, 2008a, and Rott, 2005 for similar findings). However, since the fill-in condition spent more time to complete the reading task, the beneficial effect on retention may also be attributed to this factor.

An interesting finding was that the control group reported significantly less noticing than the glossing conditions but not significantly less remembering. The think-aloud protocols revealed that a plausible explanation why remembering was higher than noticing in the control group when compared to other conditions is that participants in the control group noticed fewer words (e.g., they often skipped unknown words while reading) but noticed the word form to a greater extent than participants in the glossing conditions, and seemed to engage in deeper processing (e.g., tried to infer the word meaning, repeated the word form several times). Moreover, the remember-know task asked participants to recognize word forms and not word meanings. Therefore, these
factors may have led to reduce the difference found between the experimental and control groups in other measures with respect to the amount of remembered words. As Leow et al. (2008) and Leow et al. (forthcoming) have argued, it can be concluded from these results that there are different levels of processing at one level of awareness.

Finally, the present study supported the beneficial effect of glossing on text comprehension, as previous studies have shown (Davis, 1989; Jacob, 1994; Hee Ko, 2005; Bell and Le Blanc, 2000; Bowles, 2004; Guidi, 2009; Martínez-Fernández, 2008a; Martínez-Fernández, 2009b). Specifically, the results showed that both traditional glosses and glosses embedded in a fill-in task had a positive effect on local comprehension, that is, comprehension of the ideas expressed by the target vocabulary words. The experimental conditions turned out not to be significantly different, thus supporting previous findings (Martínez-Fernández, 2008a). In addition, types of glossing did not have any effect on global comprehension, as glossing conditions were not significantly different from the control group in their performance on global ideas. These findings led the researcher to conclude that simultaneous attention to form and meaning (i.e., lexical item and text content) did not have a detrimental effect on text comprehension or L2 learning, as previous research has suggested (cf. Han et al, 2008).

In the fourth place, the present study provided interesting findings with respect to the effect of type of linguistic item, that is, vocabulary versus grammar, on reported noticing, memory experiences, and learning. Previous literature in SLA has generally assumed that learners tend to process lexical information over grammatical information when they are exposed to the input (cf. VanPatten1994, 1996, 2004). Lexical items are
considered more salient, less abstract, and less complex than grammatical items, and thus noticing of vocabulary words is expected to be higher than noticing of grammatical items. The present study, however, indicated that the amount of reported noticing did not differ by type of linguistic item. Additionally, it was suggested that participants who reported awareness of the grammatical structure at a high level also reported awareness of a higher number of lexical items, and participants who produced fewer verbal reports of the grammatical structure reported awareness of fewer lexical items. Based on findings reported by Guidi (2009), the researcher suggested that the lack of difference for type of linguistic item in reported noticing might be explained as a result of (a) a frequency effect (i.e., multiple exposure to the grammatical structure versus single exposure to multiple lexical items) in combination with high familiarity of the verbs selected as target grammatical items, and (b) similar saliency of vocabulary words and grammatical items, given that both types of items consisted of only one morpheme. However, findings of the effect of type of linguistic item on memory experiences provided evidence in support of a different explanation. Results on memory experiences indicated that lexical items received significantly more remember and know responses than grammatical items. From this finding, the researcher concluded that the crucial difference between processing vocabulary words and grammatical structures might not lie in whether or not linguistic items are noticed but rather in subsequent processing into the language system. In other words, grammatical information may be harder to be processed further than lexical information.
Furthermore, the finding of the effect of type of linguistic item on memory experiences may explain to some extent why lexical items benefit more than grammatical items from input enhancement techniques such as traditional glosses and glosses embedded in a fill-in task. While the experimental groups were significantly different from the control group in vocabulary recognition and production measures in the present study, no significant main effect was found for condition of exposure on grammar recognition and production posttests. Thus, despite the fact that the percentage of items recognized and produced did not significantly differed by type of linguistic item, there was evidence to support some minimal vocabulary learning in the experimental groups, while the gain in grammar in the experimental groups was very low, and restricted to production of old items, with respect to the control condition. The possibility that vocabulary words are encoded in episodic and semantic memory to a greater extent than grammatical items may contribute to the effectiveness of glossing techniques in lexical development. That is, input enhancement techniques may be more efficient with items that are easier to further process into the language system than with items that are harder to further process after noticing.

In conclusion, the present study found additional support for Schmidt’s (1990) noticing hypothesis, as reported levels of awareness positively correlated with learning measures. In addition, it provided empirical evidence for the relationship between awareness and episodic memory as measured by reported experiences of remembering, and between episodic memory and learning, thus corroborating previous theoretical views in SLA (Schmidt, 1994a, Robinson, 1995b, N. Ellis, 2001, 2005). Findings did not
show robust evidence to support the effectiveness of glossing on grammar learning but did show evidence for a beneficial effect of glossing on vocabulary learning and text comprehension. However, the results did not support the Involvement Load Hypothesis proposed by Laufer and Hulstijn (2001), as the condition inducing higher involvement (i.e., glosses embedded in a fill-in task) did not differ statistically from the traditional glossing condition in reported noticing, remembering or intake of target lexical items. Finally, the present study did not find a differential effect for type of linguistic item on reported awareness, but did find a significant difference between lexical and grammatical items in reported experiences of remembering and knowing. Vocabulary words had a higher effect on remembering and knowing than grammatical items. This finding led the researcher to conclude that the difference between processing vocabulary words and grammatical structures might not lie at the stage between attention and noticing but rather at the stage between noticing and subsequent internalization of the data into the language system.

**Pedagogical Implications**

This study has several pedagogical implications. On the one hand, L1 translation glosses proved to be an efficient technique to aid text comprehension, and to a lesser extent, vocabulary learning. From the findings of this study, however, the use of glossing for complex grammatical structures should not be encouraged, as grammar learning did not appear to benefit from glossing.
On the other hand, the effectiveness of less incidental techniques such as providing glosses embedded in a fill-in task remains uncertain. While this type of glossing led to higher learning of vocabulary, this finding may be a result of simple exposure to a higher number of glossed words. In addition, this technique may cause a delay in completing the reading comprehension task. For these reasons, teachers are encouraged to use traditional glossing in combination with other methods to promote vocabulary and grammar learning.

**Limitations and Future Research**

The use of think-aloud protocols constitutes an effective method to investigate cognitive processes, as findings of the present study have revealed. Importantly, previous studies in SLA have generally found that thinking aloud while completing a task does not have a significant effect on L2 development and text comprehension (e.g., Leow & Morgan-Short, 2004; Bowles & Leow, 2005; Sachs & Suh, 2007; Bowles, 2008; Martínez-Fernández, 2008b; Medina, 2008; Yoshida, 2008; Guidi 2009). In the present study, an attempt was made to control for potential reactivity of thinking aloud. Although the statistical analyses did not indicate any significant difference between think-aloud and silent conditions, the power of these analyses turned out to be medium-low for vocabulary recognition and production, and very low for grammar recognition and production. This result indicates that a significant difference between think-aloud and silent groups cannot be completely ruled out, and a much larger sample would be necessary to confirm whether or not thinking aloud was reactive.
The present study aimed to investigate levels of awareness of both lexical and grammatical items as measured by concurrent verbalizations. However, the analysis of the think-aloud protocols did not provide evidence for reported awareness of lexical items at a high level. This result corroborated findings of previous studies (Guidi, 2009; Bowles, 2004). While previous research has examined the effects levels of awareness on L2 grammatical development (Camblor, 2006; Leow, 1997a; Rosa & O’Neill, 1999; Rosa & Leow, 2004), the effects of levels of awareness on vocabulary learning remain uncertain. Thus, this issue needs to be further investigated. In addition, reported noticing of vocabulary words in the present study did not appear to correlate with remembering, as it was expected. This finding led the researcher to suggest that remembering might correlate with higher levels of awareness. Therefore, future studies on levels of awareness of lexical items need to confirm this prediction. Moreover, results on levels of awareness of the grammatical structure in this study must be taken with caution due to the low number of participants reporting awareness at a high level.

The findings regarding the effect of two types of glossing, that is, traditional glosses versus glosses embedded in a fill-in task, on L2 development did not show enough evidence to support the Involvement Load Hypothesis. As in previous studies, the experimental conditions provided exposure to different amounts of unfamiliar vocabulary words (i.e., exposure to target words versus exposure to target and alternate words). Therefore, future studies investigating the Involvement Load Hypothesis need to control for this factor in order to confirm whether involvement (i.e., degree of evaluation) or
simple exposure to targeted and alternate words had a beneficial effect on L2 vocabulary development.

Furthermore, the glossing techniques used in the present study involved textual enhancement of the target forms plus exposure to the glosses, and consequently the control condition was provided with a text that did not include any kind of external manipulation. However, further research should control for the effect of textual enhancement by including textual enhancement in the control condition or excluding textual enhancement from all conditions. This type of research may confirm whether or not findings of the effects of glossing in the present study may be attributed to simple exposure to the glosses.

Another important limitation of the present study is that the effect of type of linguistic item was not isolated from the experimental texts that provided exposure to the target items. Different texts were used for each type of item. Although the experimental texts had similar characteristics in terms of genre, length, and author, and did not have a differential impact on global comprehension, a potential text effect must be taken into account when considering the findings. Thus, further research needs to address this limitation. Additionally, findings pertaining to the effect of type of linguistic item should be tested with a variety of structures involving different degrees of complexity. The present study suggested that the difference between processing vocabulary words and grammatical structures might be related to further processing of the linguistic data into episodic and semantic memory. However, further research examining the effect of
different types of items on noticing and episodic and semantic memory is needed in order to consolidate this finding.

Finally, long-term effects and relationships over time in the present study were examined through delayed posttests administered only one week after exposure. Thus, future studies need to explore the relationship between memory experiences, awareness and retention over longer periods of time, as well as the effects of type of glossing and type of linguistic item on retention beyond one-week-delayed posttests. Furthermore, the present study did not address experiences of recognition memory over time. Since previous cognitive psychology studies examining the effect of retention interval on memory experiences have found evidence of a remember-to-know-shift (e.g., Conway et al., 1997), SLA research needs to determine whether this shift also occurs in the context of second languages, and whether the change over time in memory experiences may have different or similar characteristics and implications as in other research areas.

To conclude, the present study has opened a new door in SLA research by investigating experiences of recognition memory in relation to L2 development, levels of awareness, input enhancement, and type of linguistic item. Future research should further investigate the role of remembering and knowing in SLA, and confirm whether or not previous findings in the cognitive psychology and neuroscience fields may be generalized to the area of second language learning.
APPENDIX A

Experimental Texts

The experimental lexical item text for the minus incidental condition

EL VERDADERO VALOR DEL ANILLO
(The real value of the ring)

Un día al salir del trabajo en el soto / muelle, un compañero le habló a Juan de un maestro que solucionaba muchos problemas. Juan le preguntó dónde vivía y decidió ir a verlo. Al llegar vio al maestro fuera de su casa, junto a un arce / una tapia. Juan le habló:

- Maestro, vengo a pedirle ayuda. Nunca he podido solazarme / descollar en nada. ¿Cómo puedo mejorar? ¿Qué puedo hacer para que la gente me aprecie más? Todos dicen que soy inútil, que no sé hacer nada bien.

Juan observó la apariencia de pordiosero / pijo del maestro y esperó. Finalmente, respondió:

- ¡Lo siento mucho, no puedo ayudarte, ¿no ves que voy a rociar / podar este árbol? Si quisieras ayudarme, quizá después podría ayudarte.

Juan observó la apariencia de pordiosero / pijo del maestro y esperó. Finalmente, respondió:

- ¿Lo sienten mucho, no puedo ayudarte, ¿no ves que voy a rociar / podar este árbol? Si quisieras ayudarme, quizá después podría ayudarte.

Juan observó la apariencia de pordiosero / pijo del maestro y esperó. Finalmente, respondió:

- Encantado, maestro. Trabajo con la tierra, soy experto en cortar las ramas de los árboles – dijo Juan, mientras observaba la camisa sucia y rota del maestro; realmente parecía un hombre muy pobre.

- Bien -dijo el maestro al terminar- Todavía necesito solucionar otro problema.

Se quitó un anillo del dedo y, dándoselo al zagal / carcamal, añadió:

- Debo vender este anillo para pagar una deuda. Vete al mercado y vénделo. No aceptes menos de una moneda de oro.

En el mercado Juan comenzó a ofrecer el anillo. Los mercaderes lo miraban con interés y después lo empezaban a elogiar / desdeñar, sobre todo al oír el precio, pues eran muy avariciosos. Solo un viejito, al ver a Juan tan joven, le explicó que ese anillo no valía tanto como él pedía. Juan regresó y le dijo al maestro con tristeza:

- Lo siento, no he podido embaucar / hallar a ningún vendedor.

- ¿Qué quieres decir?

- No es posible obtener lo que pediste. Quizás pudiera conseguir 2 ó 3 monedas de plata, pero no creo que pueda falsear el verdadero valor del anillo.

- ¡Qué importante lo que dijiste, joven! Quizá pueda ayudarnos Alfonso, el zahorí / orfebre del pueblo. Vete a verlo, dile que quieres tasar / bruñir el anillo y venderlo. Pero no importa lo que ofrezca, no se lo vendas. Vuelve con mi anillo. ¡Solo queremos saber su verdadero valor!

Alfonso examinó el anillo a la luz, lo analizó, lo pesó y dijo:

- Dile al maestro que si lo quiere vender inmediatamente, solo puedo darle 58 monedas de oro.

Juan volvió al maestro contentísimo y le contó lo que había dicho.

---

30 The font size for all texts in the original materials was Times New Roman 12pt.
Alfonso.

- Siéntate -dijo el maestro después de escucharlo- Tú eres como este anillo: una joya única y valiosa. Y como tal, solo puede evaluarte verdaderamente un experto. ¿Qué haces por la vida pretendiendo que cualquiera descubra tu verdadero valor?

Y, diciendo esto, volvió a ponerse el anillo en el dedo.

**The experimental lexical item text for the plus incidental condition**

**EL VERDADERO VALOR DEL ANILLO**

(The real value of the ring)

1. **soto**: Un día al salir del trabajo en el soto, un compañero le habló a Juan de un maestro que solucionaba muchos problemas. Juan le preguntó dónde vivía y decidió ir a verlo. Al llegar vio al maestro fuera de su casa, junto a un arce. Juan le habló:

   - Maestro, vengo a pedirle ayuda. Nunca he podido descollar en nada. ¿Cómo puedo mejorar? ¿Qué puedo hacer para que la gente me aprecie más?

2. **arce**: Juan observó la apariencia de pordiosero del maestro y esperó. Finalmente, respondió:

   - ¡Lo siento mucho, no puedo ayudarte, ¿no ves que voy a que podar este árbol? Si quisieras ayudarme, quizá después podría ayudarte.

3. **descollar**: Encantado, maestro. Trabajo con la tierra, soy experto en cortar las ramas de los árboles – dijo Juan, mientras observaba la camisa sucia y rota del maestro; realmente parecía un hombre muy pobre.

4. **pordiosero**: - Bien -dijo el maestro al terminar- Todavía necesito solucionar otro problema. Se quitó un anillo del dedo y, dándoselo al zagal, añadió:

5. **podar**: - Debo vender este anillo para pagar una deuda. Vete al mercado y vénđelo. No aceptes menos de una moneda de oro.

6. **zagal**: En el mercado Juan comenzó a ofrecer el anillo. Los mercaderes lo miraban con interés y después lo empezaban a desdénar, sobre todo al oír el precio, pues eran muy avariciosos. Solo un viejito, al ver a Juan tan joven, le explicó que ese anillo no valía tanto como él pedía. Juan regresó y le dijo al maestro con tristeza:

7. **desdénar**: - Lo siento, no he podido embauca r a ningún vendedor. ¿Qué quieres decir?

8. **embauca r**: - No es posible obtener lo que pediste. Quizás pudiera conseguir 2 ó 3 monedas de plata, pero no creo que pueda falsear el verdadero valor del anillo.

9. **orfebre**: - ¡Qué importante lo que dijiste, joven! Quizá pueda ayudarnos Alfonso, el orfebre del pueblo. Vete a verlo, dile que quieres tasar el anillo y venderlo. Pero no importa lo que ofrezca, no se lo vendas. Vuelve con mi anillo. ¡Solo queremos saber su verdadero valor!

10. **orfebre**: Alfonso examinó el anillo a la luz, lo analizó, lo pesó y dijo:

   - Dile al maestro que si lo quiere vender inmediatamente, solo puedo darle 58 monedas de oro.

   - Si, podríamos obtener por él 70 monedas, pero si la venta es urgente...

   - Juan volvió al maestro contentísimo y le contó lo que había dicho Alfonso. Siéntate -dijo el maestro después de escucharlo- Tú eres como este
anillo: una joya única y valiosa. Y como tal, solo puede evaluarte verdaderamente un experto. ¿Qué haces por la vida pretendiendo que cualquiera descubra tu verdadero valor?

Y, diciendo esto, volvió a ponerse el anillo en el dedo.

The experimental grammatical item text for the minus incidental condition

EL LEÑADOR ESFORZADO
(The hard-working woodcutter)

(2) descubrió: I found out
descubra: I find out


(3) recibió: I received
reciba: he receives

Al día siguiente a las 9 el jefe le ofreció un buen salario y Pedro se puso contento cuando recibió / reciba el contrato. - A las 11 volveré -dijo el jefe- y te daré un buen hacha para cortar árboles. Quiero ver este bosque sin ningún árbol cuando la próxima semana venga / vea el trabajo terminado.

(7) oyó: he heard
oiga: he hears

Cuando Pedro oyó / oiga estas palabras, se propuso cortar todavía más árboles. -Seré el hombre más feliz del mundo cuando mañana por la noche pueda / pude decir que he cortado 25 árboles en un día- le dijo a su mujer. A la mañana siguiente Pedro se levantó temprano y fue al bosque. A pesar del esfuerzo, no pudo cortar más de quince árboles. “Debo estar cansado. Seguro que recuperaré fuerzas cuando duerma / dormí la próxima noche y cortaré más árboles”, pensó.

(9) duerma: I slept
dormí: I slept

Al amanecer, se levantó decidido a superar su récord. Sin embargo, esa mañana consiguió cortar muy pocos árboles. “Tendré más energía cuando dentro de una hora comí / coma algo”, pensó. Pero al final del día solo había cortado ocho árboles. Al día siguiente fueron siete árboles, luego cinco y cuando el domingo repitió / repitió su jornada de trabajo solo pudo cortar un árbol. “El jefe pensará que no soy un buen leñador cuando mañana le decía / diga que sólo he cortado un árbol, y no me dará más trabajo”, pensaba Pedro, desesperado. “Iré a hablar con él cuando sale / salí de casa, a las 7 en punto”. Al día siguiente Pedro le dijo al jefe:

(12) decía: I told
diga: I tell
decía / diga que sólo he cortado un árbol, y no me dará más trabajo”, pensaba Pedro, desesperado. “Iré a hablar con él cuando salga / salí de casa, a las 7 en punto”. Al día siguiente Pedro le dijo al jefe:

(14) fui: I went
vaya: I go

Solo puedo prometerle que continuará esforzándose cuando mañana haga / hice mi trabajo si usted me lo permite. El jefe le escuchó con atención y, mirando el hacha que llevaba Pedro en la mano, le preguntó:

(15) haga: I do
hice: I did

—¿Cuándo afligaste el hacha por última vez, Pedro?
— Mmmm… —el leñador se quedó pensando, sorprendido— No tuve
tiempo, señor, estuve demasiado ocupado cortando árboles.

*The experimental grammatical item text for the plus incidental condition*

**EL LEÑADOR ESFORZADO**

(El trabajo duro del leñador)

Un día Pedro el leñador salió a buscar trabajo. Cuando volvió a casa, su mujer le preguntó qué tal había ido.


Al día siguiente a las 9 el jefe le ofreció un buen salario y Pedro se puso contento cuando recibió el contrato.

- A las 11 volveré -dijo el jefe- y te daré un buen hacha para cortar árboles. Quiero ver este bosque sin ningún árbol cuando la próxima semana venga a ver tu trabajo.

- No se preocupe, señor, empezaré a trabajar cuando tenga el hacha, a las 11 en punto, y me esforzaré al máximo —dijo Pedro.

A las 11 Pedro fue al bosque a cortar árboles. En un solo día cortó dieciocho árboles y por la noche se lo dijo al jefe.

- ¡Felicidades, continúe así! Te pagaré más cuando el próximo lunes vea el trabajo terminado.

- Gracias, señor.

Cuando Pedro oyó estas palabras, se propuso cortar todavía más árboles.

- Seré el hombre más feliz del mundo cuando la próxima noche duerma, pensó.

Al amanecer, se levantó decidido a superar su récord. Sin embargo, esa mañana consiguió cortar muy pocos árboles. “Tendré más energía cuando dentro de una hora coma algo”, pensó. Pero al final del día solo había cortado ocho árboles. Al día siguiente fueron siete árboles, luego cinco y cuando el domingo repitió su jornada de trabajo solo pudo cortar un árbol.

“El jefe pensará que no soy un buen leñador cuando mañana diga que sólo he cortado un árbol, y no me dará más trabajo”, pensaba Pedro, desesperado. “Iré a hablar con él cuando salga de casa, a las 7 en punto”. El día siguiente Pedro se dijo al jefe:

- Señor, ayer cuando fui al bosque solo pude cortar un árbol. Me he esforzado hasta el límite, no sé qué me pasa, cada día me esfuerzo más y más y cada día corto menos árboles. Le aseguro que yo soy un buen leñador. Solo puedo prometerle que continuaré esforzándome cuando mañana haga mi trabajo si usted me lo permite.

El jefe le escuchó con atención y, mirando el hacha que llevaba Pedro en la mano, le preguntó:

- ¿Cuándo afilaste el hacha por última vez, Pedro?
- Mmmm... —el leñador se quedó pensando, sorprendido— No tuve tiempo, señor, estuve demasiado ocupado cortando árboles.

---

(1) volvió: he came back
(2) descubra: find out
(3) recibió: he received
(4) venga: I come
(5) tenga: I have
(6) vea: I see
(7) oyó: he heard
(8) pueda: I am able to
(9) duerma: I sleep
(10) coma: I eat
(11) repitió: he repeated
(12) diga: I tell
(13) salga: I leave
(14) fui: I went
(15) haga: I do
APPENDIX B

Instructions for five groups (a, b, c, d, and e)

(a) Instructions for the [-I, +TA] group

Please read the following story at your pace. As you read, you will find pairs of words in bolding face. The meanings of these words are provided in the margins. You need to highlight in yellow the one that you think it fits better in the context. Please pay attention to the words in the margins, and don’t try to guess. You can go back to change your options as you continue reading. In addition, you will have to think aloud all your thoughts to the recorder. Whenever something comes to your mind, please try to say that aloud. You can use any language at any point while you think aloud.

After reading, you will have to answer a few questions in English about the content of the passage. Please write down the time when you start to read and when you finish, and don’t forget to read the title of the story.

(b) Instructions for the [-I, -TA] group

Please read the following story at your pace. As you read, you will find pairs of words in bolding face. The meanings of these words are provided in the margins. You need to highlight in yellow the one that you think it fits better in the context. Please pay attention to the words in the margins, and don’t try to guess. You can go back to change your options as you continue reading.

After reading, you will have to answer a few questions in English about the content of the passage. Please write down the time when you start to read and when you finish, and don’t forget to read the title of the story.

(c) Instructions for the [+I, +TA] group

Please read the following story at your pace. As you read, pay attention to the words in bolding face. The meanings of these words are provided in the margins. In addition, you will have to think aloud all your thoughts to the recorder. Whenever something comes to your mind, please try to say that aloud. You can use any language at any point while you think aloud.

After reading, you will have to answer a few questions in English about the content of the passage. Please write down the time when you start to read and when you finish, and don’t forget to read the title of the story.
(d) Instructions for the [+I, -TA] group

Please read the following story at your pace. As you read, pay attention to the words in bolding face. The meanings of these words are provided in the margins.

After reading, you will have to answer a few questions in English about the content of the passage. Please write down the time when you start to read and when you finish, and don’t forget to read the title of the story.

(e) Instructions for the [++I, -TA] group

Please read the following story at your pace. After reading, you will have to answer a few questions in English about the content of the passage. Please write down the time when you start to read and when you finish, and don’t forget to read the title of the story.
APPENDIX C

Text for Think-Aloud Practice

EL VIAJE DE LOS ZAPP

Dicen que la vida es un viaje improvisado y para la familia argentina Zapp, que ha recorrido todo el continente americano a bordo de un auto de 1928, los miles de kilómetros recorridos nunca son demasiados si se trata de realizar un sueño.

Herman y Candelaria Zapp comenzaron su aventura en enero del año 2000 con la intención de viajar desde Argentina hasta Alaska. El plan inicial era partir a pie, pero al final iniciaron el camino en un coche con ruedas de madera y un motor de cincuenta y cinco caballos que circulaba a una media de cincuenta kilómetros por hora.

Querían llegar a Alaska en seis meses para luego volver a Argentina y tener hijos, pero cuando llegó el tiempo límite aún estaban en Ecuador construyendo un pequeño barco, con la ayuda de decenas de indígenas, que les permitiera cruzar el Amazonas hasta Brasil. Ese momento fue uno de los peores del viaje, pues no tenían dinero ni para gasolina.

Candelaria empezó a pintar cuadros sencillos con pájaros de los lugares que iban conociendo y Herman aprendió a enmarcar. Con el dinero de las ventas siguieron avanzando. Tardaron cuatro años en alcanzar su destino y además lo harían con un hijo nacido en California y su coche convertido en cocina, salón y dormitorio en la parte superior.

Como en todo viaje, los imprevistos sorprendieron a la familia Zapp: se quedaron sin dinero, se les averió el coche y hasta les denegaron los visados. Sin embargo, consideran que ese momento de crisis económica se ha convertido en la mejor época de sus vidas.

En su camino de regreso a Argentina tuvieron otro hijo y al llegar a casa escribieron un libro, Atrapa tu sueño, donde narran sus aventuras y con cujos beneficios de ventas piensan costearse su futuro viaje en Asia y, por qué no, poco a poco, dar la vuelta al mundo.

(Texto adaptado de www.diariovasco.com)
APPENDIX D

Reading Comprehension Pretest

Text

Directions. Please read the text at your pace. Don’t forget to read the title of the story. After reading, you will have to answer a few questions in English about the content of the passage.

EL CÍRCULO DEL 99
El cuento de las 99 monedas de oro
(The Circle 99. The 99 gold coins story)

Había una vez un rey muy triste que tenía un sirviente, llamado Hasán, que era muy feliz. Un día el rey le preguntó cuál era el secreto de su felicidad. Hasán le explicó que no tenía motivos para estar triste: tenía su mujer y sus hijos viviendo en la casa que les había asignado y el rey les daba ropa y comida, ¿cómo estar triste? El rey estaba furioso, ¡esa no podía ser la razón de su felicidad! El rey le pidió una explicación a su consejero y este le explicó que Hasán era feliz porque estaba fuera del círculo del 99. Como el rey seguía sin comprender, el consejero decidió demostrarle lo que era.

Por la noche el rey y su consejero fueron hasta la casa de Hasán. El consejero dejó en su puerta una bolsa con 99 monedas de oro, llamó y fue a esconderse con el rey detrás de un árbol. Hasán abrió la puerta y vio la bolsa, miró hacia todos lados y entró en la casa. No podía creerlo. ¡Era una montaña de monedas de oro! Empezó a agrupar las monedas de 10 en 10 y las fue sumando, 10, 20, 30… hasta que formó el último montón… ¡9 monedas! Buscó en la mesa, en el piso, movió todos los muebles, pero no encontró nada. "No puede ser, ¡me la robaron! Noventa y nueve monedas es mucho dinero. Pero 99 no es un número completo. ¡Cien es un número completo!".

Hasán puso la bolsa debajo de la cama y empezó a pensar. ¿Cuánto tiempo tendría que ahorrar dinero para comprar su moneda número 100? Hizo el cálculo. En doce años juntaría lo necesario. "Doce años es mucho tiempo. Pero mi esposa también puede buscar trabajo en el pueblo", pensó. Él mismo trabajaría en el palacio hasta la noche para recibir alguna paga extra. Quizá así en siete años juntaría el dinero. ¡Pero era demasiado tiempo! Siguió pensando. Vendería un poco de comida para obtener más dinero. Y vendería también su ropa de invierno. En realidad, tenían demasiada ropa de invierno y hacía mucho calor… "Sí, es un sacrificio, pero después todo será diferente".

El rey y el consejero observaron la cara de preocupación de Hasán y volvieron al palacio. El consejero le explicó al rey que Hasán había entrado en el círculo del 99. Durante los siguientes meses, Hasán siguió sus planes. Una mañana, el sirviente tenía que hacer la cama del rey y entró a su habitación golpeando las puertas, de mal humor.

REY: ¿Hasán, qué te pasa?
HASÁN: Nada.
REY: Antes, no hace mucho, reías y cantabas todo el tiempo.
HASÁN: Hago mi trabajo, ¿no? ¿Qué cree que soy, Majestad? ¿Su bufón?

Al poco tiempo el rey decidió expulsar al sirviente. No era agradable tener un sirviente que estaba siempre de mal humor.
Questions

Directions. Please answer in English the following questions about the content of the passage. If you don't know or don’t remember an answer, just write "NO".

1. Please explain what the king asked Hasán at the beginning of the story, what Hasán answered, and how did the king reacted to Hasán’s answer.
2. What did the king’s advisor explain to the king?
3. Why did the king and his advisor go to Hasán’s place that night?
4. Where did the king and his advisor hide themselves to observe Hasán?
5. Why did Hasán get crazy when he counted the coins in the bag?
6. Hasán thought that there were 99 coins instead of 100 because… (complete the sentence)
7. Where did Hasan hide the bag?
8. Please list all things that Hasán planned to do to get the missing coin.
9. What happened when Hasán entered the king’s room?
10. In your opinion, what does “getting into the circle 99 mean?”
APPENDIX E

Reading Comprehension Questionnaires

Reading Comprehension Questionnaire for the Lexical Item Text

Directions. In this task, you need to (a) provide short answers in English to questions about the content of the passage you just read, or (b) complete sentences containing a blank based on the content of the passage you just read. If you don't know or don't remember an answer, just write "NO"

1. Where did Juan work?
2. How did Juan hear of the master?
3. Why was this master well-known?
4. When Juan saw the master for the first time, he was standing outside his house, next to a __________
5. Juan went to see the master to ask him for help. Juan told the master that he had never __________
6. According to Juan, what does people think about him?
7. Juan observed the master, who looked like a __________
8. The master told Juan that he could not help him in that moment because first he had to __________
9. The master asked Juan to help him out with two tasks. Please indicate what the master asked him.
10. The master took off his ring, and gave it to the __________
11. In the market, the merchants looked at the ring with interest, and then they started __________ the ring, especially when they were told the price.
12. Please explain what happened in the market. Did Juan meet his goal? Why?
13. When Juan comes back from the market, he says to the master: “I am sorry, I have not been able to __________ anyone.”
14. Why does Juan think that he could not sell the ring?
15. The master thinks that Alfonso might be able to help them. Alfonso is a __________
16. The master wants Alfonso to __________ the ring.
17. What did Alfonso tell Juan?
18. Did Juan sell the ring to Alfonso? Why?
19. What did the master tell Juan when Juan told him the news?
20. In your opinion, what is the meaning of this story?
Reading Comprehension Questionnaire for the Grammatical Item Text

Directions. In this task, you need to provide short answers in English to the following questions about the content of the passage you just read. If you don't know or don’t remember an answer, just write "NO"

1. What did Pedro say to his wife when he found a job?
2. Why did Pedro accept the job?
3. What did the boss say to Pedro when he told him how many trees he had cut down?
4. What did Pedro think it would make him the happiest man in the world?
5. What happened on the second day when Pedro went to work?
6. Why was Pedro afraid after his third day working?
7. Why did Pedro think that he could not cut down so many trees as he cut down at the beginning?
8. When Pedro explained to his boss what it was happening to him, what did the boss ask him?
9. What did Pedro answer?
10. In your opinion, what is the meaning of this story?
APPENDIX F

The Remember-Know Task

Instructions

Have you seen the following words in ANY of the passages that you read in the second session (i.e., *The woodcutter* and *The real value of the ring*)? In this task, you are going to see 50 words. When you see a word you will have to provide the following information:

A) Was this word in any of the texts? Yes / No
B) If Yes: do you *remember* the word in the texts, or do you *know* that the word was in there, or are you *guessing*?

In order to distinguish *remember/know/guess* responses, and understand what we mean by these words, please read carefully the following definitions:

- **Remember**: This is an experience characterized by the specific recollection of contextual details involving one’s self at a particular time and in a particular place. For example, if I tell you the title of a movie you might be able to remember when exactly you watched that movie, with whom, how you were feeling that day, and other contextual details. In the same way, when you have read a text, and afterwards you encounter a word that was in that text, you might be able to “see” the word in the text again, that is, to remember exactly where it was in the text (at the top or at the bottom of the page, in the middle, at the right or left margin…), or how it appeared (bolded, italic, etc.), or you might remember what you thought when you saw that word, or you might remember yourself focusing on the word, reading it aloud, etc. If you have any of these experiences when you see any of following words, please select “YES-Remember.”

- **Know**: This is an experience characterized by a more abstract awareness of familiarity. For example, this is the experience that you have when you know that you have seen a movie, but you are not able to remember when, where, with whom, how you were feeling that day, etc. In the same way, you might know that a specific word was in a texts that you just read, and not be able to remember yourself reading the word, thinking of the word, or remember where on the page the word was, how it appeared, etc. If you have this experience when you see any of the following words, please select “YES-Know.”

- **Guess**: Finally, you might try to guess whether a word was in the texts that you read or not (when you are *not able to remember* the word and you *don’t know* if the word was in the text). For example, you might select “Yes, it was in the text” because the meaning of the word could fit in the story. If this is the case when you see any of the following words, please select “YES-Guess”.

- **No**: If you know that the word was not in any of the two texts you read, please select “No.”
In addition, you will have to **think aloud** while you complete this task, and explain why you are selecting one answer or the other. Please say aloud any specific memory you have (for example, the context of the word, in what text it appeared, where on the page it appeared, if you remember yourself thinking of that word, etc.). If you have at least one specific memory, you can select “remember.”

*Example*

Orfebre

a. YES-Remember
b. YES-Know
c. YES-Guess
d. NO

*Items*

1. ganar
2. descubra
3. soto
4. elogiar
5. veo
6. agricultor
7. descollar
8. duerma
9. calcular
10. orfebre
11. vendré
12. rociar
13. embaucar
14. salga
15. monte
16. pordiosero
17. podrá
18. desdeñar
19. pijo
20. tenga
21. zahorí
22. tasar
23. dormiré
24. cultivar
25. zagal
26. descubriré
27. tapia
28. podar
29. vea
30. carcamal
31. arce
32. salgo
33. rechazar
34. niño
35. venga
36. mentir
37. bruñir
38. pueda
39. planta
40. tengo
41. hallar
42. coma
43. digo
44. muelle
45. haga
46. comeré
47. solazarse
48. hago
49. diga
50. joyero
APPENDIX G

Recognition Tests

_Meaning Word Recognition Test_

Note. Correct answers are marked in bold. Both target and alternate words are shown.

Directions. In this test, each question presents a word in Spanish. Please select the answer that provides the translation of the word. **DON'T TRY TO GUESS.** If you don't know the answer, just select the "I don't know" option.

1. tasar
   a. to value
   b. I don't know
   c. to scorn
   d. to taste

2. rociar
   a. to polish
   b. **to sprinkle water**
   c. I don't know
   d. to rise

3. soto
   a. maple tree
   b. **grove**
   c. I don't know
   d. shout

4. zahorí
   a. **future-teller**
   b. beggar
   c. I don't know
   d. zealot

5. pordiosero
   a. **beggar**
   b. young man
c. port authority
d. I don't know

6. muelle

a. landing pier
b. mule
c. I don't know
d. fence

7. podar

a. I don't know
b. to value
c. to prune
d. to be able to

8. tapia

a. tap
b. fence
c. I don't know
d. landing pier

9. orfebre

a. goldsmith
b. young man
c. saffran
d. I don't know

10. bruñir

a. to praise
b. to brush
c. I don't know
d. to polish

d. to be outstanding
12. pijo
   a. old man
   b. pigment
   c. posh person
   d. I don't know

13. orfebre
   a. I don't know
   b. orphan
   c. goldsmith
   d. posh person

14. carcamal
   a. I don't know
   b. old man
   c. caramel
   d. future-teller

15. embaucar
   a. to be outstanding
   b. to embark
   c. to trick
   d. I don't know

16. solazarse
   a. I don't know
   b. to solicit
   c. to have fun
   d. to sprinkle water

17. hallar
   a. to hallow
   b. to have fun
   c. I don't know
   d. to find
18. arce
   a. grove
   b. I don't know
   c. arch
   d. maple tree

19. desdeñar
   a. I don't know
   b. to trick
   c. to descend
   d. to scorn

20. elogiar
   a. to allege
   b. to find
   c. to praise
   d. I don't know

Grammar Recognition Test

Note. Correct answers are marked in bold. Target items (old and new items) are shown.\textsuperscript{31}

Directions. In this test, you will read 30 sentences containing a blank. Please select the answer that fits best in the sentence in each case. \textbf{DON'T TRY TO GUESS}. If you don't know the answer, just select the "I don't know" option

1. Empezaré a trabajar cuando en unas horas ________ el hacha y me esforzaré al máximo.
   a. tengo
   b. I don't know
   c. tenga
   d. tendré

2. El maestro se sorprenderá cuando en unos minutos le ________ la nueva noticia sobre el verdadero valor del anillo.
   a. traiga
   b. I don't know

\textsuperscript{31} In the pretest, the sentences had the same structure as in the posttests but names of the characters and objects were replaced with others.
3. Te pagaré más cuando el próximo lunes ________ el trabajo terminado.
   a. veo  
   b. vea 
   c. I don't know  
   d. veré

4. Me sentiré mejor mañana, cuando _________ la ayuda del maestro.
   a. recibiré  
   b. I don't know  
   c. recibo  
   d. reciba

5. Estoy seguro de que aprenderé mucho cuando _________ su consejo.
   a. oiga  
   b. I don't know  
   c. oigo 
   d. oiré

6. Decidiré si quiero este trabajo cuando mañana _________ las condiciones. Veré al jefe a las 9:00.
   a. descubro  
   b. descubra  
   c. descubriré  
   d. I don't know

7. Solo puedo prometerle que continuaré esforzándome cuando mañana _________ mi trabajo si usted me lo permite.
   a. haga  
   b. I don't know  
   c. hago  
   d. haré
8. Recordaré las palabras del maestro cuando la próxima vez ________ que la gente no me escucha.

   a. sienta
   b. I don't know
   c. sentiré
   d. siento

9. Contaré la historia del maestro cuando ________ un libro el próximo año.

   a. escribo
   b. escribiré
   c. escriba
   d. I don't know

10. Quiero ver este bosque sin ningún árbol cuando ________ la próxima semana.

    a. I don't know
    b. vengo
    c. venga
    d. vendré

11. No sé si el maestro me ayudará cuando mañana yo le ________ ayuda.

    a. I don't know
    b. pida
    c. pediré
    d. pido

12. Iré a hablar con él cuando mañana ________ de casa, a las 7:00 en punto.

    a. saldré
    b. I don't know
    c. salgo
    d. salga

13. A partir de ahora la gente me escuchará con atención cuando en el próximo encuentro yo ________ las palabras del maestro.

    a. I don't know
    b. repetiré
    c. repita
    d. repito
14. Seguro que recuperaré fuerzas cuando _________ la próxima noche y cortaré más árboles.

   a. duermo  
   b. duerma  
   c. dormiré  
   d. I don't know

15. Le traeré al maestro el dinero cuando _________ del mercado, sobre las 6:00 de la tarde,  
además de vender el anillo.

   a. volveré  
   b. I don't know  
   c. vuelvo  
   d. vuelva

16. Seré el hombre más feliz del mundo cuando mañana por la noche _________ decir que he  
cortado 25 árboles en un día.

   a. podré  
   b. pueda  
   c. puedo  
   d. I don't know

17. Mañana le contaré mi problema al maestro cuando _________ a verlo.

   a. vaya  
   b. voy  
   c. iré  
   d. I don't know

18. Tendré más energía cuando _________ algo dentro de una hora.

   a. como  
   b. comeré  
   c. I don't know  
   d. coma

19. El jefe pensará que no soy un buen leñador cuando mañana le _________ que sólo he cortado  
un árbol.

   a. digo  
   b. diga  
   c. diré  
   d. I don't know
20. El maestro estará contento cuando yo ________ su anillo en las próximas horas.

a. venda
b. I don't know
c. vendo
d. venderé
APPENDIX H

Written Production Tests

Meaning Word Production Test

Note. This test includes the same items as the meaning word recognition test (see Appendix G).

Directions. In this test, each question presents a Spanish word. Please provide an English translation or definition of each word. If you don't know the meaning of a word, just write "NO"

Example.

1. arce: ________________________________________________________________
2. embaucar: __________________________________________________________

Grammar Production Test

Note. This test includes the same items as the grammar recognition test (see Appendix G).

Directions. In this test, each question shows a sentence containing a blank. Please fill in the blank by writing down in the answer box below the correct form of the verb provided in parentheses. If you don't know what form you need, just write down "NO"

Example.

1. No sé si el maestro me ayudará cuando mañana yo le ________ (pedir) ayuda.
2. Tendré más energía cuando ________ (comer) algo dentro de una hora.
APPENDIX I

Post-debriefing Questionnaire

1. Gender:
   1. male
   2. female

2. What is/are your native language/s?

3. If you know any language other than Spanish, English or your native language, please specify which one and what your proficiency level in that language is (beginner, intermediate, advanced).

4. Please rate how much you liked the text "The hard-working woodcutter"
   1. I liked it a lot
   2. I liked it
   3. Indifferent
   4. I didn't like it
   5. I didn't liked it at all

5. Please rate how difficult to understand "The hard-working woodcutter" was.
   1. Very difficult
   2. Difficult
   3. Neither difficult nor easy
   4. Easy
   5. Very easy

6. Please rate how much you liked the text "The real value of the ring"
   1. I liked it a lot
   2. I liked it
   3. Indifferent
   4. didn't like it
   5. I didn't liked it at all

7. Please rate how difficult to understand "The real value of the ring" was.

8. Have you encountered, looked up or learned any of the following words outside the experiment (i.e., prior to the first session or between the second and the third session)?

   SOTO, MUELLE, TAPIA, ARCE, PIJO, PORDIOSERO, ORFEBRE, ZAHORÍ, ZAGAL, CARCAMAL, DESCOLLAR, SOLAZARSE, DESDEÑAR, ELOGIAR, BRUÑIR, TASAR, EMBAUCHAR, HALLAR, PODAR, ROCIAR
If so, please write down which ones, and explain where or how you encountered or learned them.

9. Have you looked up outside the experiment any specific grammatical structure that encountered during the experiment?

10. Were you asked to think aloud while reading the texts? ___ If so, was it difficult to think aloud while reading? ___________ If yes, please explain why.

11. Was it difficult to think aloud while making Remember/Know/Guess judgments? If so, please explain why.

12. Was it difficult to distinguish a "remember experience" from a "know experience"? (Please note that the question is not whether it was easy to remember the words or not, but just how easy or difficult to make the judgment was). If it was difficult, please explain why.

13. Was it difficult to distinguish a "know experience" from a "guess experience"? If it was difficult, please explain why.

14. Did you notice any frequent grammatical forms in the text "The woodcutter" WHILE YOU WERE READING IT (not after)? If yes, please specify which ones.

15. Did you figure out any grammatical rule while reading "The hardworking woodcutter"? If so, please try to explain the rule that you picked up while reading.

16. Do you have any suggestions or comments about this study?


Pineros (Eds.), *Theory, Practice, and Acquisition: Papers from the 6th Hispanic Linguistics Symposium and the 5th Conference on the Acquisition of Spanish And Portuguese* (pp. 395-411). Somerville, MA: Cascadilla.


Schooler (Eds.). *Scientific approaches to the question of consciousness* (pp. 13-47). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.


University of Pittsburgh.


