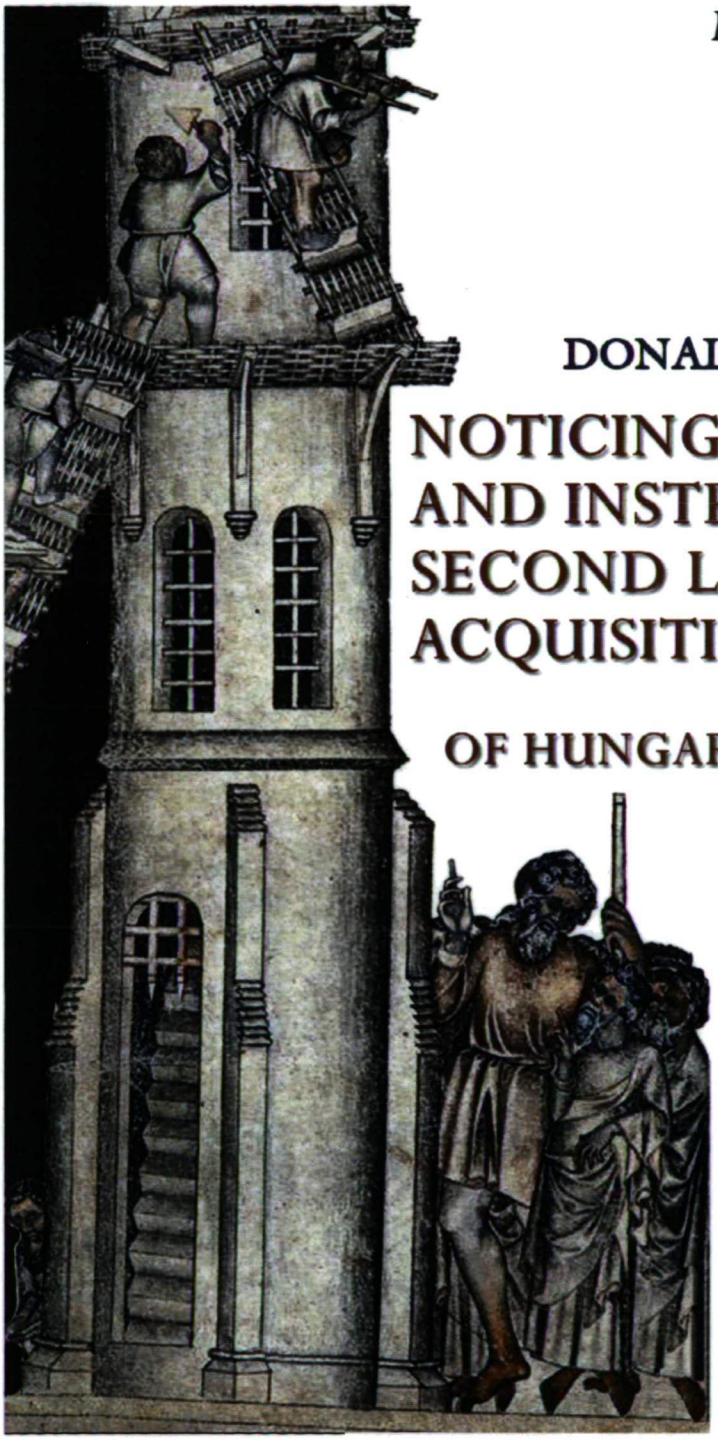
PAPERS IN ENGLISH & AMERICAN STUDIES XVI.
Monograph Series 6.



DONALD W. PECKHAM

NOTICING AND INSTRUCTION IN SECOND LANGUAGE ACQUISITION: A STUDY OF HUNGARIAN LEARNERS OF ENGLISH

JATEPress
SZEGEDI EGYETEMI KIADÓ
Szeged 2009

Papers in English & American Studies XVI

Papers in English & American Studies XVI.
Monograph Series 6.

DONALD W. PECKHAM

NOTICING AND INSTRUCTION IN SECOND LANGUAGE ACQUISITION:

A STUDY OF HUNGARIAN LEARNERS OF ENGLISH

JATEPress
Szeged 2009

Papers in English & American Studies is published by the
Institute of English & American Studies (IEAS)
of the University of Szeged
H–6722 SZEGED, HUNGARY
Egyetem u. 2.
<ieas@lit.u-szeged.hu><www.arts.u-szeged.hu/ieas>

The production of the present volume was sponsored by
the Textbook Fund of the University of Szeged

General editor:
ZOLTÁN VAJDA
the Director of IEAS

Publisher's Reader:
MARIANNE NIKOLOV

Designer:
ETELKA SZŐNYI
The cover illustration was designed on the basis of "Babel",
a 14th c. illumination of Egerton Genesis folio 5v

ISSN 0230–2780
ISBN 978–963–482–946–1

CONTENTS

Preface 7

Chapter 1. Introduction 9

Chapter 2. Review of the literature 11
2.1. Background on the question of the role of consciousness in language learning 11
2.2. Interface positions 12
2.3. Refinement of the definition of consciousness and attention as applied to SLA 13
2.4. The role of noticing in second language learning 15
2.5. The promotion of noticing through instruction 16
2.6. A memory-based framework for accounting for noticing 18
2.7. Noticing grammar vs. noticing vocabulary 21
2.8. Summary and implications 22

Chapter 3. Hypotheses 23

Chapter 4. Methodology 25
4.1. Participants and setting 25
4.2. Target grammar and vocabulary 29
4.2.1. Grammar 30
4.2.2. Vocabulary 36
4.3. Design and outline of procedure 37
4.4. Instruction and exposure 38
4.4.1. Reading texts and reading tasks 39
4.4.2. Grammar and vocabulary instruction 41
4.4.3. Exposure to grammar and vocabulary 43
4.5. Instrumentation 44
4.5.1. The noticing test 44
4.5.2. Grammar tests 50
4.5.3. Vocabulary tests 51
4.6. Scoring and analysis 52
4.6.1. Scoring the noticing test 52
4.6.2. Scoring the grammar tests 53
4.6.3. Scoring the vocabulary tests 53
4.6.4. Adjusting grammar and vocabulary scores for previous knowledge 54
4.6.5. Analysis of the retrospections 54

Chapter 5. Results and discussion for noticing 55
5.1. Overview of results concerning noticing 56
5.2. Results concerning the effects of instruction vs. exposure on the noticing of grammar 58
5.3. Results concerning the effects of instruction vs. exposure on the noticing of vocabulary 60
5.4. Results concerning the noticing of instructed target item grammar vs. non-instructed grammar 61
5.5. Results concerning the noticing of instructed target item vocabulary vs. non-instructed vocabulary 62
5.6. Results of the comparison of target grammar items which were the object of exposure with target vocabulary items which were the object of exposure 63
5.7. Discussion 64
5.7.1. Hypotheses 1 and 2 64
5.7.2. Hypotheses 3 and 4 70
5.7.3. Hypothesis 5 72
5.7.4. Results using Remember + Know responses to define noticing 73
5.8. Summary 74

Chapter 6. Results and discussion concerning noticing and learning 77
6.1. Results for learning 77
6.2. Results plotting noticing against learning 79
6.3. Discussion of Hypotheses 6 and 7 85
6.4. Further interpretation of noticing and learning data 88
6.4.1. Analysis and discussion of high and low learning and noticing scores 88
6.4.2. Further discussion of correlational data 91
6.5. Results when using Remember + Know scores to calculate noticing 92
6.6. Summary concerning a relationship between noticing and learning 94

Chapter 7. Individual differences and individual variation in noticing 95
7.1. Consistency of noticing scores across time shown by correlations 95
7.2. Consistency of scores shown by individuals 97
7.3. Summary and conclusions 105

Chapter 8. The noticing of individual target grammar and vocabulary items 107
8.1. Overview of results on individual items 107
8.2. Individual items which show the largest differences between groups 113
8.3. Individual items which show the highest scores overall 117
8.4. Comparison of accuracy orders for individual items between groups 121
8.5. Summary 123

Chapter 9. General discussion 125
9.1. Noticing grammar and vocabulary in input 125
9.1.1. The limits of the impact of instruction on noticing 126
9.1.2. The differential effect of instruction on grammar and vocabulary noticing 128
9.1.3. Evaluation of responses to the noticing test 130
9.2. The relationship between noticing and learning 134
9.3. Conclusions concerning the noticing test itself 135

Chapter 10. Conclusions and implications 137

APPENDICES 139
Appendix A: An example of the instructional materials used in the study 139
Round One, lesson 1, grammar instruction 139
Round One, lesson 1, vocabulary instruction 143
Appendix B: Instructions for the noticing test 146
Appendix C: And example of texts used in the noticing test 148
Round One, test 1, text for noticing 148

REFERENCES 149

PREFACE

In the field of second language acquisition (SLA), cognitive theories of language learning and the research based on them have received growing interest and attention since the 1990s, and the research I present in this monograph fits squarely within this tradition. This project, which investigates the relationship between foreign language instruction and the noticing of features of language in input, was originally written as my PhD dissertation at the University of Pittsburgh, and was carried out at a Hungarian secondary school. Since the completion of this research, the field of second language acquisition has continued to explore cognitive approaches to language learning, confirming the importance of the role of explicit processes such as noticing as studied in this research.

As will be developed in the following pages, noticing refers to the conscious, explicit registration of linguistic form in working memory, and has been seen as an essential step in the process of learning, as first presented in the "noticing hypothesis" (Schmidt, 1990). Noticing can be conceived of on the most basic level of attending to forms and functions in input or on the more complex level of consciousness raising concerning forms and their functions, but in either case, it is this intersection with conscious awareness and input which is essential for learning. This point of view, that consciousness is key to learning, has stood in contrast to vastly popular approaches to second language learning which promoted a view where second language acquisition happens as an essentially unconscious, implicit process (e.g. Krashen, 1981; 1985). By now it is safe to say that while both implicit and explicit processes exist in language learning, there is a quite large role for explicit processes and there is no reason to assume that implicit processes alone are responsible for learning. The noticing of form in input, then, becomes a potentially vital first step in the learning of grammar and vocabulary, and the research presented here in this monograph can add to the understanding of that process.

The research developed here contributes to the understanding of noticing through the establishment of links between instruction and noticing, and through the development of a theoretically grounded instrument for the study of noticing. A main claim of the study is that foreign language instruction indeed induces learners to notice grammar and vocabulary in subsequent input. Furthermore, a differential relationship was found in the data between the noticing of grammar and vocabulary, whereby vocabulary is easier to notice, yet instruction in grammar produces greater results in terms of noticing. These findings, though, rest on the ability of a research instrument which allows learners to express their conscious awareness of form in input. To this end, a test of noticing was developed based in part on work in recognition memory where the differential levels of conscious awareness of previously encountered data has been researched (see, for example, Gardiner, Ramponi, & Richardson-Klavehn, 1998). The resulting "noticing test" developed for this current research allows learners to report on whether forms encountered in input were accompanied with conscious recollection

(that is, noticing) or not. This principled test of noticing has potential to be applied in other contexts as well.

The role of noticing continues to be an important topic both inside and outside of the classroom. Since 2000, when this research was completed, there has been a growing amount of research and theoretical work concerning a cognitive approach to language learning, establishing clear links between conscious processes, learning and instruction. (See, for example, Robinson, 2001, Doughty, 2003, and DeKeyser 2007). Furthermore, the links between interaction and learning, which have been long established, have been put on a cognitive footing whereby attention and noticing play a key role highlighting those features of input which are used by learners for the development of their second language system (see Gass, 2003 and Gass & Selinker, 2008, chapter 10, for current reviews). Indeed the relationship between interaction and language learning has received tremendous emphasis in recent years as the context of language learning and use has been greatly emphasized through the development of the concept of multicompetence (see Hall, Cheng, and Carlson, 2006). Inherent in these approaches is the view that it is through learners' interaction in specific social contexts that language is learned. From the perspective of a cognitive approach to second language learning, it is important to understand how consciousness and the noticing of forms and functions develop and play a role in learning through interaction in specific contexts.

I hope that the availability of the research presented in this monograph can contribute to the work of others on the important topics of noticing and second language learning, and to research on cognition and language learning in general. The conclusions which follow from this research have practical applications to language learning (Peckham, 2001), and the noticing test developed here has potential for applications in the further study of explicit processes in language learning.

Finally, I would like to thank Robert DeKeyser for his invaluable assistance in completing this research, Nikolov Marianne for her generous help with reviewing this manuscript, and especially my family in Hungary for their constant and continued support of me throughout this work.

Szeged, Hungary, December 2008

CHAPTER 1. INTRODUCTION

The general issue that this project addresses is how second language learners use and apply explicit, conscious knowledge to develop and extend their competence in their second language. While this might seem to be an issue which should have already been thoroughly researched, over the past decades the complexities of this problem have often been largely ignored as practical and theoretical perspectives, some of which appeal strongly to what seems to be common sense, have held sway. Thus the common sense perspective appealing to the "accumulated entities" or building blocks model of the acquisition of grammatical form, where each grammatical structure is added to the learner's language one block at a time, oversimplifies the use of explicit knowledge and ignores the complexities involved in language learning in general (Lightbown, 1984). On the other hand, what might be described as an overextension of the clear thinking brought about by the re-evaluation of the simplistic building block model through the application of a much more sophisticated model of general competence — communicative competence — and the development of a much more sophisticated and subtle model of linguistic competence — exemplified by the transformational generative model and its direct successors — eventually led to the development of second language learning theories and teaching practice such as the theory of creative construction (Dulay & Burt, 1975) and the practice of the strong version of the communicative approach to language teaching, which denied or ignored any large role that learners' explicit knowledge has in the development of competence.

At the present stage of the development of various theories of second language acquisition (SLA), cognitive approaches to the problem of second language learning have been formally developed which view the issues of the role of explicit and implicit knowledge in a relatively sophisticated way (see Skehan, 1998, for a general overview and Johnson, 1996, for a specific approach). Using these models, researching the role of explicit knowledge in second language learning does not represent the pendulum swing back to old views, but a spiraling up to using more sophisticated approaches in order to deal with a vital issue.

The specific questions that this study addresses are, first, whether classroom instruction in specific grammatical forms and vocabulary influences learners to consciously notice the items they were taught once they encounter them in subsequent input through reading. And secondarily, the question of whether there is a relationship between the noticing of items and their learning is addressed. These issues are important ones as strong claims have been made concerning the necessity of noticing items in input for language learning (Schmidt, 1990), and also weaker claims have been made which possibly provide a large role in the process of learning for noticing features of input (e.g. R. Ellis, 1994). Also, as another issue in this project, the question is raised whether grammar and vocabulary differentially require instruction for

them to be noticed in input. This problem is also followed up in terms of which items of vocabulary and grammar appear, through the data collected here, to be more susceptible to noticing. Also, individual differences among participants concerning noticing are discussed.

Data for this project were collected in a bilingual, Hungarian-English, secondary school in Hungary. Participants were first-year students at the high school and were at an intermediate or upper intermediate level. Existing classes were used, and the specially created instructional materials were integrated, as much as possible, into the curriculum of the school. The research design involved dividing participants into two groups whose attention was focused either on particular grammatical constructions or vocabulary items. After two instructional sessions, participants were given an immediate and two delayed post-tests. These tests consisted of a noticing test which presented participants with a text and then measured their noticing of specific grammar and vocabulary items, and a test of learning for those same items. Classroom-based research of this kind is important if a role for the use of explicit knowledge in language learning is to be tracked and tested in natural, real-life settings.

One of the main issues faced in designing the research that investigates what participants notice is the creation of a subtle enough test which will allow for participants to express very fine distinctions concerning consciousness of grammar and vocabulary seen in input. Thus, for this study I utilized a framework for researching recognition memory, the so-called "Remember/Know" paradigm, which allows participants, after training, to report rather subtle states of consciousness. As well as providing a vehicle for collecting responses from participants concerning their memories, this framework relates quite nicely to the theoretical framework which is used to describe and explain noticing as developed in the SLA literature. Previously, noticing has been tested in a variety of ways including the collection of on-line verbal protocols, and thus one of the contributions of this study is in the development and testing of one more alternative way of measuring participants' consciousness of grammar and vocabulary which they have experienced in input.

CHAPTER 2.
REVIEW OF THE LITERATURE

2.1. BACKGROUND ON THE QUESTION OF THE ROLE OF CONSCIOUSNESS IN LANGUAGE LEARNING

One of the most heated debates in the area of SLA in the past decades has been over the role of consciousness in the language learning process: are consciousness of linguistic structures and conscious knowledge about those structures facilitative in learning a second language, or are they, at best, stop gap measures which can be used only in the right circumstances but which do not contribute to the development of interlanguage? This question and ones like it were eventually proposed in the light of the re-evaluation and rejection of then current behaviorist theories which had been extended to language pedagogy and a theory of language learning in the late 1950s and early 1960s. Since that time and up until recently, theoretical positions on language learning have either focused on areas of learning where consciousness is quite unlikely to be a relevant issue, such as questions concerning the role of Universal Grammar in SLA, or have been dominated by views flatly rejecting any positive role of consciousness in contributing to L2 (second language) acquisition (e.g. Krashen, 1981, 1985).

Krashen's claim, which has had tremendous influence, was that consciously learned knowledge could not be used in building competence, but only knowledge that was acquired naturally through comprehensible input could become competence. Now, it is currently accepted that the "acquisition vs. learning" distinction is more productively thought of in terms of the implicit-explicit distinction (Hulstijn & Schmidt, 1994). Thus, the question can then be put in terms of whether there is an interface between explicit knowledge about language and the implicit knowledge that underlies fluent and automatized language production. Krashen's "no interface" position accepts the idea that explicit knowledge can lead to consciously controlled language use, but sees this system as separate from the system of implicit knowledge that competence is based on. In this view, implicit knowledge is developed through interaction with input with no conscious focus on form.

For nearly a decade starting from the mid 1970s, Krashen's theories supporting a no-interface view dominated the field of SLA, due in part to an effort to identify in L2 acquisition those same implicit and highly modularized processes and mechanisms which were being researched in L1 (first language) acquisition, and also due to the perceived failure of traditional language teaching methods which stressed explicit rule learning and mechanical drilling. Soon after that, in the early 1980s, a growing number of dissenting voices emerged which promoted models of SLA in which there was a role for consciousness (Bialystok, 1982; Long, 1981; Sharwood-Smith, 1981). Research on the role of consciousness has actually benefited from the long debate with the no-interface position, not by claiming a decisive win in the

debate — nor by coming up with a unified position on the issue — but by being faced with the requirement of looking to more sophisticated models of cognition and memory in which to couch a theory of SLA.

For this reason, much current SLA research taps into models of attention, memory, and consciousness which are found in cognitive psychology. Now, more than a decade later, the climate has changed to the point where SLA researchers are no longer forced to defend the very idea that consciousness and explicit knowledge have a role in language learning, but can begin with an assumption that an interface position — in some form — is a valid starting point for research (Hulstijn & De Graaff, 1994:101).

Indeed, as it has been pointed out, it is actually rather unusual that so much effort over the past years has been spent trying to show that explicit learning and explicit instruction does not play a role in second language learning, while in cognitive psychology, where explicit knowledge has always been considered to play a central role in learning, it is the issue of implicit knowledge and implicit learning which has been questioned more thoroughly (DeKeyser, 1994). Furthermore, even Reber, one of the earliest and strongest advocates of the possibility of the implicit learning of abstract patterns, in an article co-authored with Winter clearly states concerning the evidence for implicit learning that "this is not to suggest, of course, that embedded rule systems cannot be grasped within awareness — no one could argue that conscious analytical strategies do not play a fundamental role in learning" (1994: 118).

2.2. INTERFACE POSITIONS

Views claiming a positive relationship between explicit and implicit knowledge in language learning can be summarized as the strong and weak interface positions. The strong-interface position (e.g. Bialystok, 1978; DeKeyser, 1998; Sharwood-Smith, 1981) supports the view that explicit knowledge can be used in the development of implicit knowledge through practice and the resulting automatization that occurs in this process. The strong interface view relies on models of skill acquisition developed in cognitive psychology. Anderson's model of skill acquisition (1983) has been applied to several areas, for example language transfer (Faerch & Kasper, 1987), language learning strategies (O'Malley & Chamot, 1990) and second language classroom learning in general (DeKeyser, 1998). Anderson's model proposes a three stage process of learning where declarative knowledge is first learned and manipulated slowly, then proceduralized in a process where subroutines of the skill are chunked together into higher-order commands. Finally, the skill is fine tuned and automatized, resulting in fluent language use based on procedural, and most likely implicit, knowledge. Although this model would suggest that skilled and automatic performance of language always begins with the use of explicit rules, Chamot & O'Malley (1994) point out that the learning of procedural knowledge can also conceivably be done through the observation and imitation of "the expert performance of a model who has a meaningful message and has a valued communication goal",

or through the practice of the components of a skill while receiving feedback on performance (p. 397).

The weak-interface point of view (e.g. R. Ellis, 1990, 1993) sees no direct connection between explicit and implicit knowledge, but accepts that explicit knowledge can influence the creation of implicit knowledge through making forms salient. Thus, according to R. Ellis (1993), explicit knowledge can aid learners in noticing elements in input, in noticing the gap between the input and the learner's current state of interlanguage, and in monitoring before and after output. Finally, following views on learning and developmental orders (e.g. Pienemann, 1984), R. Ellis argues that if learners are at the appropriate developmental level, explicit knowledge can be directly converted into implicit knowledge. Thus, unless the learner is at the proper stage for learning a particular grammatical form, explicit knowledge can only have a delayed impact on learning.

Recently, a growing number of studies supporting an interface position and showing the positive benefits of attention and explicit knowledge in learning have been carried out from a variety of perspectives and in a variety of situations. Alanen (1995) studied the acquisition of semi-artificial Finnish morphology, showing a benefit for rule-based instruction over control groups, yet also finding an overall effect for the benefits of attention on learning. Rosa & O'Neill (1999) looked at the effect of attention on intake, finding that attention to and noticing of forms in input had a large effect on intake and that task demands produce different effects on attention paid to form. Izumi, Bigelow, Fujiwara, and Fearnow (1999) studied the effects of attention to form in output and were able to show some effects of focusing learners' attention on output in incorporating new grammatical forms. Finally, comparing explicit and implicit learning in laboratory-like conditions, DeKeyser (1995) demonstrated a stronger effect for explicit learning of simpler rules and no difference between implicit and explicit learning of complex rules.

2.3. REFINEMENT OF THE DEFINITION OF CONSCIOUSNESS AND ATTENTION AS APPLIED TO SLA

A view that supports the use of explicit knowledge in second language learning must involve a clear definition of how it is that learners attend to and interact with explicit knowledge in learning. Thus, recently various models and explanations of attention and consciousness in language learning have been proposed.

It is now accepted that consciousness is not a unitary phenomenon in SLA, but is multifaceted and can be seen with reference to a variety of different processes and constructs. Schmidt (1994) breaks down the concept of consciousness into four different levels: consciousness as intentionality (intentional vs. incidental learning); consciousness as attention (focal attention vs. peripheral attention); consciousness as awareness (explicit vs. implicit learning); and consciousness as control (controlled vs. automatic processing). Of central in-

terest for this study is the concept of "noticing", which, according to Schmidt, is a phenomenon which occurs at the level of consciousness as attention.

Schmidt refers to noticing as the "conscious registration of the contents of focal attention" (1994:17), that is, the subjective experience of a particular formal feature of linguistic input. It is important at the outset here to distinguish this level of consciousness implied by "noticing" from the level of consciousness implied by "awareness" as represented in the levels of consciousness presented above. On this view, "awareness" refers to higher level awareness of a rule or generalization — what is also sometimes referred to as "understanding" (Schmidt, 1995:29) — while "noticing" refers to the conscious registration of some feature of the input. While these divisions certainly cannot create water-tight distinctions, they do serve to define noticing as a lower-level phenomenon which does not necessarily imply explicit knowledge of a rule or generalization, but should not exclude noticing as a result of attempting to identify rules and generalizations.

As implied by the framework above, noticing can be located in a model of attention, in particular, Tomlin & Villa's model of the role of attention in SLA (Tomlin & Villa, 1994). They divide attention into four components: detection, alertness, orientation, and awareness. According to their view, of these four parts, detection of stimuli and relationships between stimuli is the one necessary element which allows for further processing of input. Here, detection refers to "the process by which particular exemplars are registered in memory" and thus are made available for further processing (p. 193). Alertness — the readiness to receive input — and orientation — the direction of attentional resources towards some specific aspect of the input — can both increase the likelihood of detection, but are not required. Finally, "awareness" in their model — what is defined as consciousness at the level of noticing, in Schmidt's taxonomy — can positively affect alertness and orientation, but again, is not required. Schmidt's definition of noticing as the conscious registration of input can be accounted for in Tomlin & Villa's model as the presence of all four elements: awareness of what is detected (noticing), along with the concurrent alertness and orientation required for awareness. Tomlin & Villa see detection as sufficient for the further processing and possible acquisition of features of the input, while Schmidt claims that noticing is "the necessary and sufficient condition for the conversion of input to intake for learning, on the grounds that all demonstrations of detection without conscious registration (blind-sight, subliminal perception) demonstrate only the processing of what is already known, not learning," (Schmidt, 1994: 17).

This rather strong statement on the part of Schmidt represents what he calls the "noticing hypothesis". It essentially states that there is no learning without conscious awareness at the level of noticing. This position echoes established points of view in cognitive psychology which suggest that awareness in terms of what is being learned and the ability to report on that learning are generally the default case (Ericsson & Simon, 1993), and that claims of learning without awareness can often be effectively challenged due to a lack of sensitivity to what participants are actually aware of (Shanks & St. John, 1994) This should not, though, rule out "incidental learning" — the learning of something which is outside of *focal* attention (Hulstijn, 1989; 1992) — nor does it rule out the possibility of implicit learning of *some* kinds of material (Carr & Curran, 1994; DeKeyser, 1995). That is, implicit learning of local co-occur-

rence patterns may be possible, although the implicit learning of more abstract and long distance patterns may not. It should also be noted here that explicit learning can play a larger or smaller role in learning depending on the types of structures that are being learned. Prototypicality patterns of morphology, for example, are less likely to be positively affected by explicit learning than the learning of non-contiguous, long-distance relationships (DeKeyser, 1995; Hulstijn & De Graaff, 1994). Furthermore, concerning the *requirement* of the presence of noticing for learning, Schmidt himself, referring to Baars (1988), states that "given the fact that it may be impossible to agree upon an operational definition of noticing that will allow falsifiability of this hypothesis, it may be wiser to replace zero-point claims (no learning without noticing) with a modified hypothesis that more noticing leads to more learning" (1994:17-18). Baars states in more detail:

> Unfortunately in the case of learning, most discussion of the role of consciousness seems to be assimilated to the "necessary condition" question. But even if conscious experience were not a necessary condition but only a helpful adjunct to the learning process, it would be difficult to doubt that in the real word consciousness and learning are very close companions. Thus the controversy about the *necessity* of consciousness tends to interfere with a more subtle question about the role consciousness plays in *most* cases of learning. (1988: 217)

What this suggests, then, is that the noticing hypothesis can be most profitably seen as a view of learning which strongly supports the facilitative effects of consciousness on second language learning. This should not imply that the "zero-point" position is necessarily untenable — indeed, in a recent review of literature on attention, memory, and SLA (Robinson, 1995b) this position was largely upheld. But the strongest view of the noticing hypothesis does remain controversial, as was pointed out in another review (Truscott, 1998) which raises questions about the viability of noticing more abstract and complex grammatical structures. The purpose of this research project, though, is not to answer this question of the zero-point interpretation of noticing, but to look at the impact of instruction on noticing and the subsequent potential facilitative aspects of that consciousness produced.

2.4. THE ROLE OF NOTICING IN SECOND LANGUAGE LEARNING

Noticing plays a key role in various theories of SLA as a means of converting input into "intake". Intake itself is defined variously in different theories. The term was coined by Corder (1967) to describe what "goes in" as opposed to all of the input that is available to go in — that is, the difference between the input that is merely processed and the input which is used in the development of the linguistic system. Since that time, the definition of the term has also emphasized the process which goes on between input and the integration of rules in the learners'

interlanguage (Chaudron, 1985). In this way, Gass (1988: 207) defines intake as "the process of assimilating linguistic material". For her, the result of this process can either be storage of an element in memory or the development of the second language rule system, with consciousness playing a role in this process of storage and integration. Beyond these differences between a product or process focus, what is common in these definitions is the view that intake is involved in a mediating process which has the potential to integrate into the linguistic system some of the linguistic material represented.

Various models have been put forth for the role that noticing plays in converting input into intake. As was noted above, Schmidt supports the view that noticing is "necessary and sufficient" for converting input into intake and suggests that there is no second language learning without noticing. Evidence for Schmidt's view is found in his diary study documenting his learning of Portuguese over a several-month period (Schmidt & Frota, 1986). According to the study, before elements were incorporated into Schmidt's production they first needed to be noticed in the input. Elements which may have been present in the input for months were thus not actively used until they were first noticed, according to Schmidt. Gass' model slightly predates the current discussion of noticing, but also appears to assign a crucial role to noticing in the process of intake and, ultimately, integration. Noticing is seen by Gass, in an early formulation, as "apperceived input" (1988). The term is borrowed from psychology, where it is used to refer to perception which is influenced by past perceptions, that is, "a cognitive act which identifies that form as being related to some prior knowledge which has been stored in our experience" (p. 200–201). She goes on further to note that apperception can be seen as a "priming device" for use in analyzing input — that is, input which has been noticed before has a greater potential to be noticed and analyzed in the future. R. Ellis (1993) assigns noticing a weaker role than Schmidt, supporting the weak interface position, where explicit knowledge can indirectly aid in the development of implicit knowledge through the highlighting of form in input. Ellis describes this process as "intake facilitation". Thus, for R. Ellis, noticing input facilitates learning, while for Schmidt, noticing is essential for learning. In each view, though, despite disagreements about the ultimate necessity of noticing for learning, noticing potentially plays a key role in the conversion of input into intake, which is eventually integrated into the interlanguage system.

2.5. THE PROMOTION OF NOTICING THROUGH INSTRUCTION

The question of consciousness is relevant to both naturalistic and instructed L2 learning, though it is particularly relevant for instructed learning. Noticing can be cued by a variety of factors which are often referred to in the literature: frequency of occurrence in the data; novelty; prior knowledge; and various other events which would draw attention to a particular form, such as problems in communication or the specific task demands of the situation. One of the more important questions that has to do with classroom-based learning is the role of explicit instruction in facilitating the development of the learner's grammatical system. One

of the claims for the efficacy of explicit instruction — in various forms — is that it can make input salient for the learner (Doughty, 1991; R. Ellis, 1993; Sharwood-Smith, 1991, 1993). Putting this in terms of noticing, the claim is that explicit instruction is valuable in that it leads to the noticing of features of the input, which, in turn, can possibly lead to intake and the development of the learner's grammatical system.

One of the earliest studies which looked at the effect that instruction has on noticing as defined here was done by Fotos (1992). She carried out a study to compare the effects of two kinds of instruction — "formal teacher fronted lessons" and problem solving "consciousness raising activities" — on drawing attention to grammatical form, specifically, indirect object placement, adverb placement, and relative clause usage in English as a second language. One week after instruction, participants were given tests to determine which grammatical elements were noticed in input. Two types of tests were given, each one consisting of a task carried out while listening to a text — one, a listening comprehension task, the other a dictation — each of which were followed by directions telling students to look at a printed version of the text and underline any "special uses of English" that were present. Results compared with a control group, who did not receive instruction, showed that both instructed groups scored significantly better than the control group (which scored extremely low or zero on most tests). Furthermore, there were no significant differences between the two instructed groups, showing that both types of instruction had a similar effect. Proficiency gain tests showed that both instructed groups improved over the course of instruction. Finally, the correlation between noticing and proficiency gains was not significant, except for the teacher-fronted instruction group, which showed a significant correlation for only one grammatical form. These conclusions suggest that explicit instruction leads to more noticing when compared to a control group, though the connection between noticing and acquisition is unclear.

Fotos' study showed in a very broad way that explicit instruction can lead to increased noticing of grammatical forms. What the study lacked, and what is needed for looking at the question of noticing in general, is a convincing psychological framework in which to place noticing and the instruments used to measure it. Couching noticing in a theory of memory and cognition can give greater precision to the concept of noticing, and will possibly allow for more fine-grained measurement. Further studies since that time have used more sophisticated frameworks to address the issue of noticing.

A study by Jourdenais, Ota, Stauffer, Boyson, and Doughty (1995) investigated the issue of whether the enhancement of texts through highlighting target forms would increase the noticing of those forms. Two groups of beginning learners of Spanish were given texts containing the preterit and imperfect verb forms. For one group the verb forms were highlighted using textual enhancement, and for the other no enhancement was used. In a later narrative writing task participants were instructed to think aloud concerning the construction of their texts. Analysis of these verbal protocols showed that the group which received the enhanced texts made more reference to the target forms, that is, noticed more, and also used more of the target forms in the texts which they constructed when compared to the control group.

Two studies by Leow (1997, 1998) have looked in detail at the issue of attention and second language acquisition. In both studies, beginning learners of Spanish were exposed to

irregular preterit forms of Spanish verbs under a variety of conditions which promoted different levels of awareness of the structures. Results showed that participants who showed more awareness of the forms of the target structures showed greater gains than less aware participants in recognition and production tasks following instruction. Following Tomlin & Villa's (1994) framework, Leow (1998) showed that participants scored higher if they attended to the forms at the level of detection than at the levels of alertness and orientation. Detection in this case was operationalized as participants commenting on the form on-line during a problem solving task, and/or as participants correcting an error that they made with the form during the problem solving task.

In a study by Rosa & O'Neill (1999), the awareness of Spanish conditionals was manipulated through four different experimental groups and a control group which varied in the dimensions of +/– instruction, +/– directions to search for rules. Awareness was defined as the ability to make a verbal report concerning the target items while completing a task requiring the use of those items. Intake was then measured through a multiple-choice recognition task. Results showed that noticing has a large effect on intake in that the greater awareness there was, the greater the intake that occurred.

These studies in general show a positive effect for instruction on noticing and provide some indication that noticing has a positive effect on learning. Only one of the studies reviewed here, though, Fotos (1992) investigated these issues in an actual classroom situation, and this study was the least sophisticated in terms of its theoretical orientation towards the issues of attention and noticing. This is precisely one area, using classroom-based research, where this present research project can contribute to the current research being done on the issue of noticing.

A second area where this project can add to the research on noticing is through developing and implementing a new method of measuring noticing. Noticing has been measured through a variety of means: underlining noticed forms in texts (Fotos, 1992; Izumi et al., 1999), simply responding to a question as to whether something was noticed or not (Robinson, 1995a), and the use of on-line verbal reports, and think aloud protocols while completing a task involving the target items (Leow, 1997, 1998; Rosa & O'Neill, 1999). This study will introduce yet a different method of measuring noticing based on techniques used in recognition memory research. It is this issue which will be taken up below.

2.6. A MEMORY-BASED FRAMEWORK FOR ACCOUNTING FOR NOTICING

As was noted in section 2.3, there is a conflict between Schmidt's view that detection and noticing (consciousness) are required for learning, and Tomlin & Villa's view that detection alone is required. Robinson, in his review article on attention, memory, and the noticing hypothesis, sees a resolution between these two points of view by defining noticing as "detection plus rehearsal in short-term memory, prior to encoding in long-term memory" (1994b: 296). Consciousness then emerges after a certain level of activation is achieved in short-term

memory. Although detection can occur without awareness, the result of this brief, subliminal exposure "cannot be in any useful sense be claimed to be evidence of learning" (p. 298). Conscious awareness, in Robinson's view, is the result of rehearsal and elaboration, and is essentially a by-product of task demands which encourage processing likely to engage awareness. Data-driven processing — that is, processing which results in the gradual strengthening of frequency patterns — is less likely to result in noticing, while conceptually driven processing — more elaborative top-down processing involving the activation of schemas, etc. — is more likely to result in noticing. Schmidt (1995), though, disagrees with this processing view and sees consciousness as a determining factor in learning, noting that the correlation between learning and awareness is "too high to be coincidental" (p. 28), that is, not just a by-product of processing. Whatever the conclusion of the argument over the nature of consciousness, the value of this discussion is to place noticing within the larger framework of the research on memory and attention. For this reason, in this study I will be using Robinson's (1995b) definition of noticing as detection coupled with a degree of rehearsal in short-term, or working memory. Indeed, current conceptions of working memory see it as a place where recollection of previous events can be compared and integrated with current experiences and input (Baddeley, 1993).

At this point it is important to emphasize that noticing does not necessarily refer only to the first occurrence of noticing of the feature in the input. That is, noticing should not necessarily be equated with "discovering" features of the input, but can be an event which is repeated over time, or which occurs with greater or lesser strength, depending on task dimensions, type of processing or other features. This is important because the claim of the impact of instruction on noticing is that instruction enhances the later noticing of the instructed features in the input. From the subjective point of view of the learner, then, noticing can be defined as a conscious detection of features of the input which either results in the "discovery" of the features or the strengthening and enhancement of the memories of features already noticed. This can be the result of the effects of novelty, the accumulation of frequency effects, contextual saliency, the priming effect of instruction, or even the conscious search for features or rules previously encountered in instruction or elsewhere. One need not be conscious of the source of the experience which led to noticing, only the effects of noticing: conscious registration — however fleeting — of a feature of the input.

The methodological problem, though, is how to measure a subjective event such as consciousness. That is, if a person notices something in the input, how is it possible to verify at a later time that a person was consciously aware of something about those features? Here, research on consciousness and memory can provide some direction.

Tulving (1985) relates three different types of consciousness to three different types of memory: procedural memory, semantic memory, and episodic memory. Procedural memory is memory for how skills are carried out, and corresponds with "anoetic" consciousness, the ability to react to immediate environmental stimuli. Semantic memory is memory for knowledge about the world, and corresponds to "noetic" consciousness, the ability to be aware of and act on this symbolic knowledge. Episodic memory — most important for this study — is memory for "personally experienced events", and corresponds to "autonoetic" conscious-

ness, consciousness that "confers the special phenomenal flavour to the remembering of past events" (p. 3). Tulving's claim is that these different types of memory necessarily co-occur with the corresponding types of consciousness. The implication of this is that the three pairs of different types of memory and their corresponding types of consciousness can be separated from each other. Thus, it is possible to have knowledge that we don't know the exact source of — say, what Americans typically eat for breakfast — and knowledge that we possess due to personal experience — say, what we personally ate for breakfast this morning.

Based on differences between episodic and semantic memory, a corresponding distinction can be drawn between "remembering" and "knowing". Thus, according to Tulving (1985):

> If it is possible to recover knowledge about past events from either the episodic system or the semantic system, then the phenomenal experience that accompanies the recovery of such information may be one of remembering (autonoetic awareness) or knowing (noetic awareness), or a mixture of the two. It follows, then, that one way of measuring autonoetic awareness could take the form of asking people, when they recall or recognize a previously encountered item, whether they *remember* the event or whether they *know* in some other way that it occurred. The probability of the 'remember' judgement can serve as an index of the extent to which autonoetic consciousness is involved in recovery of knowledge about past events in a particular situation. (p. 6)

Thus, remembering necessarily entails a memory of consciously experiencing that event when the information was encountered, while knowing entails no autonoetic consciousness of the event when the information occurred. It is also important to note here that in terms of memory, consciousness can refer to being conscious of the act of retrieval and/or the product of the retrieval process (Schacter, 1989). When talking about Remember responses, "consciousness" is clearly referring to the product, that is, the memory of consciousness at the time of study.

This distinction between remembering and knowing has been empirically tested by Tulving and others. The typical experimental procedure involves presenting participants with a list of words to memorize which they are later tested on by deciding whether words seen on a subsequent list were present in the first list or not. Participants mark each item they perceive as being from the previous list as either an item they remember from the list — that is, they recall the event of experiencing the word on the list (for whatever reason) — or know the item was on the list — that is, they are sure the item was on the list, but have no recollective experience of it having been there. Although most all research using the Remember/Know distinction has been based on the learning of word lists, one recent study (Conway, Gardiner, Perfect, Anderson, and Cohen, 1997) has used the distinction to look at the knowledge students have after attending academic classes. Concerning word list learning, clear dissociations between Remember and Know responses have been shown concerning the effects of depth of processing, generation effects, divided attention, word and non-word recognition, and length of retention (Gardiner, 1988; Gardiner & Java, 1990, 1991; Gardiner & Parkin, 1990). These effects follow the same patterns that are found in dissociations between explicit

and implicit memory, and thus despite the fact that recognition tests are explicit tests of memory (Richardson-Klavehn & Bjork, 1988), Gardiner & Parkin (1990) see Know responses similar to cued recall tests as reflecting a type of implicit memory. Furthermore, these results can be seen to support the hypothesis that remembering reflects more elaborative and conceptually-driven processing, while knowing shows data-driven processing (Gardiner & Parkin, 1990). This is consistent with Robinson's memory-based view of noticing in SLA, reviewed above, in which noticing emerges in short-term memory out of elaborative and conceptually driven processing.

The Remember vs. Know distinction can thus be interpreted within the framework of noticing as it has been developed in SLA. If noticing refers to the conscious experience of a particular feature of the input, then a later test which includes this particular feature will potentially elicit a Remember response if that conscious event occurred with sufficient strength. Experience with input which did not result in consciousness of a particular feature — or resulted in low-level, fleeting consciousness of the feature — has the potential to result in a Know response on a subsequent test. All of this, furthermore, occurs within a context of the demands that particular tasks make on learners, and the kinds of elaboration, processing, or rehearsal that result from these demands. Divided attention, depth of processing, and other related factors should thus directly affect what is noticed, as should other factors such as perceptual salience of features and previous exposure to the features.

A test of memory involving the Remember/Know distinction can be classified and described in the following way. Gardiner & Java (1993) suggest that it is appropriate to describe tests of memory in terms of the type of testing task, the type of consciousness involved and the type of memory system implicated. Thus, the test for noticing used in this research is an explicit test of memory which relies on the states of consciousness of Remembering and Knowing produced by the two long-term memory systems of episodic and semantic memory.

2.7. NOTICING GRAMMAR VS. NOTICING VOCABULARY

One of the questions which is important concerning explicit instruction is in which domains and for what kinds of structures explicit instruction is effective (Hulstijn & De Graaff, 1994). This same question can be asked in terms of noticing by considering the types of items in input that students are most likely to notice due to the effects of instruction. It is quite possible that grammar, being more abstract, may be more difficult to notice than vocabulary items, which exist on a more "surface" level. For this reason, it is important to look at both grammar and vocabulary separately in this study. While predictions concerning the differential effects of instruction on grammar and vocabulary are difficult to make, it is possible to predict that due to the abstract nature of grammar, it will be less likely to be noticed through exposure than vocabulary, which is much more tangible for learners.

The distinction between vocabulary and grammar being made here in this project can be seen to reflect the differences between "item" learning and "rule" learning. Concerning this

distinction, Hulstijn & De Graaff suggest the hypothesis that rule learning will be more likely to benefit from instruction than item learning due to the fact that items are more salient to learners (1994:105). Although the distinction between items and rules is not necessarily parallel with that of the distinction between vocabulary and grammar, the distinction does support the differentiation between more surface level, salient elements of language, and more abstract, less salient elements and the view that the way these items are learned may be different. Again, extending this to noticing, this suggests the possibility that grammar and vocabulary may be differentially noticed. One way of investigating this is to compare the noticing of grammar and vocabulary that participants are exposed to in order to see if, outside of instruction, one domain is easier to notice than the other.

2.8. SUMMARY AND IMPLICATIONS

Noticing appears to be possible in situations which allow for elaborative processing of input. Instruction can play a role in noticing by making features of the input salient through focusing learners' attention on specific aspects of form, and even by providing contextual information that may facilitate conceptually-driven processing. Furthermore, specific task demands can influence the processing of input. That is, even with the same input, learners can potentially notice different elements of that input based on the demands placed on them and the instructions given to them. Thus, explicit instruction can have a direct effect on noticing through drawing attention to forms by providing situations which promote appropriate processing. What this suggests, then, is that different ways in which learners encounter grammar and vocabulary in the classroom — and in natural communication — will have varying amounts of potential for activating noticing when these same structures are encountered again. Finally, if models of SLA are correct which assign a role to noticing, those structures which are noticed in input should stand a greater chance of becoming intake and being integrated into the developing grammatical system. This, then, provides a link between instruction, noticing, and interlanguage development.

CHAPTER 3. HYPOTHESES

From the preceding review of the literature, the following seven hypotheses emerge, which can be grouped into four general categories. Some explanation and elaboration will accompany each category.

A. Central hypothesis: instruction vs. systematic exposure

Hypotheses 1 and 2 concern the effects of instruction vs. systematic exposure on the noticing of grammatical structures and vocabulary in input. As will be defined later, "systematic exposure" refers to those items which participants were exposed to while their attention was focused elsewhere. These hypotheses make between-group predictions, and they are the central hypotheses of this study.

Hypothesis 1

Participants who receive instruction in grammar will notice those items in input to a greater extent than those participants who have merely been exposed to those items in previous input.

Hypothesis 2

Participants who receive instruction in vocabulary will notice those items in input to a greater extent than those participants who have merely been exposed to those items in previous input.

B. Instruction vs. incidental exposure

Hypotheses 3 and 4 make within-groups predictions and involve the comparison of items which were the target of instruction and those items which were not instructed but simply were present in the text used in the test of noticing, that is those items which received "incidental exposure". Evidence for these hypotheses would provide support for Hypotheses 1 and 2. If instructed grammar and vocabulary are easier to notice than items which only incidentally appeared in the tests for noticing, it would show that noticing is a product of instruction, rather than simply an artifact of the test used for noticing.

Hypothesis 3

Grammatical structures which were the object of instruction will be noticed to a greater degree than grammatical structures which were not the object of instruction, but which appeared in the test for noticing.

Hypothesis 4

Vocabulary items which were the object of instruction will be noticed to a greater degree than vocabulary items which were not the object of instruction, but which appeared in the test for noticing.

C. Vocabulary vs. grammar noticing through exposure

Hypothesis 5 makes a between-groups prediction, claiming that vocabulary will be easier to notice than grammar.

Hypothesis 5

Higher scores will be achieved for the noticing of vocabulary that participants were exposed to than the noticing of grammar that participants were exposed to.

D. The relationship between noticing and learning

Hypotheses 6 and 7 predict that noticing will have a positive impact on learning. These hypotheses involve the analysis of individual participant data.

Hypothesis 6

No participants will have high scores on tests of grammar learning who do not also have high scores on noticing grammar.

Hypothesis 7

No participants will have high scores on tests of vocabulary learning who do not also have high scores on noticing vocabulary.

CHAPTER 4. METHODOLOGY

This chapter will first present an overview concerning the participants of the research and the target grammar and vocabulary used. Following this, there will be an in-depth discussion of the procedures and instruments used. Since this study greatly relies on a test for noticing structures in input, the development of this instrument will be discussed in detail. Indeed, a potentially important contribution of this research project in and of itself is the development and validation of this test of noticing. This is a point which will be taken up in later chapters.

4.1. PARTICIPANTS AND SETTING

Participants for this study were 27 Hungarian secondary-school students attending an English and Hungarian bilingual high school in southern Hungary. Four existing classes of students were used in the study to form two treatment groups. Treatments were randomly designed, but as existing groups were used, this should be considered pseudo-experimental study, rather than a true experiment.

Data collection was carried out twice, in two rounds, using different target grammar and vocabulary in each round, though similar groups of participants. As data was collected twice using the same procedure and similar subjects, this can be considered a replication of the study. This was done, firstly, in order to increase the amount of data collected and available for analysis, and, secondly, to provide data concerning the consistency of the research methodology used.

Concerning the composition of the groups themselves, there were 26 participants in Round One, 13 in each treatment group, while in Round Two, there were 14 participants, 7 in each treatment group. Of the 14 participants in Round Two, 13 had also been participants in Round One, thus making the group of Round Two participants roughly a subset of Round One participants. The difference in the number of participants from Round One and Round Two is the result of the loss of two instructional sessions due to scheduling conflicts.

All participants were native speakers of Hungarian in their first year of a four-year *gimnázium*, or college-prep high-school program, and were between 14 and 15 years old. The school itself is considered one of the best of the city's eight college-prep high schools, and admission to the school is based on the results of a series of competitive exams and interviews which gauge a student's academic potential and English proficiency. Thus, the students for this study were most likely, in terms of their abilities to learn foreign languages in the classroom, above the average found in Hungary. Furthermore, the four teachers described their students as generally motivated and serious.

The curriculum of the high school provides students with content area instruction in English for all subjects except Hungarian history, language and literature, which students take throughout the four years. Most content area courses are reported to be taught in a traditional teacher-fronted manner, though students do experience some classes in which more student involvement and participation is possible. Along with academic subjects taught in English, students are also required to take English as a Foreign Language (EFL) courses which are held in all four years. In the first year, these courses are centered around a typical EFL course book, while in the upper grades the focus is more specifically on academic language skills.

Most of the teachers at the school are Hungarians who completed a five-year university degree program, usually double majoring in English and another subject such as history or another foreign language. Native speaker teachers also work at the school, usually visiting on exchange programs from secondary schools in the United States or through other arrangements; the school has a policy of providing a native speaker teacher for students in at least one class a day.

As first year students, the participants in this study attended 15 hours of subject area courses in English per week, along with an additional six hours of EFL instruction. Four of these six hours of EFL instruction were spent using an up-to-date EFL text with supplementary materials provided by the instructors. The text used in all classes was the revised edition of the upper-intermediate level text from the *Headway* series (Soars & Soars, 1998). This text is an integrated four skills text, and while grammar and vocabulary instruction is systematically integrated into and presented in the text, the outward focus is on skill development and communication. Instructors supplemented the text with a stronger focus-on-form, bringing in additional exercises and systematically leading students through a rather comprehensive review of grammar which incorporated all of the forms found in an upper-intermediate level text. The remaining two hours of EFL instruction a week were spent in communication classes with native speakers. These classes were not intended to provide any overt focus on form, but instead gave students a chance to participate in discussions and communicative activities.

Overall, then, through academic classes in English as well as EFL classes, participants in the study received exposure to English in a variety of situations from listening to lectures on academic subjects, to formal grammar instruction and participation in communicative activities. Outside of the language input received in school, participants reported an interest in watching English language TV (available via the local cable service) and using the internet in English. None of the participants reported regularly speaking English with family or friends outside of the classroom.

At the time of the study, all participants in both groups had completed their first semester at the high school and were considered by their teachers to be at an "intermediate" level, similar to the level of having passed the Cambridge First Certificate Exam. This is roughly borne out through participants' scores on a grammar proficiency exam and from performance on reading comprehension exercises. To assess overall grammatical proficiency, I gave all the participants the shortened version of the Oxford Placement Exam (Allen, 1992). From a total possible score of 50 points, participants in Round One scored 37.26 points on average, and participants in Round Two scored 38.57 points. According to the information provided by the author of the exam, these scores roughly correspond to the intermediate, or Cambridge

First Certificate, level, that is, CEFR level B2. Furthermore, concerning the assessment of the participants' reading ability, during the course of the study, participants were asked to read modified passages and answer comprehension questions taken from the reading comprehension component of mock First Certificate exams contained in exam preparation guides (Hashemi, 1991; University of Cambridge Local Examinations Syndicate, 1990). From a total possible of 16 points, participants in Round One scored 14.03 points on average, while participants in Round Two scored 13.86 points. These are rather high scores, possibly indicating the proficiency level reflected in having passed the First Certificate Exam.

From a pool of 39 students in the four classes, data from 27 students were eventually used in either one or both of the two rounds of data collection. Participants' data were eliminated for the following reasons: 1) having spent more than a year in English speaking countries; 2) missing both instructional sessions; or 3) missing one or more testing sessions. In the end, data from 26 students were used in the first round of the study, and data from 14 were used in the second. Of the 14 participants from Round Two, data from 13 were used in Round One as well, as noted above.

Table 4.1. General group characteristics for Round One and Round Two for both experimental groups.

Groups	Proficiency scores			Years studying
	Grammar	Vocabulary	Reading	English
Round One				
Gram. instruction/ vocab. exposure				
M	37.39	28.77	13.85	6.85
SD	5.02	4.34	1.41	1.52
Vocab. instruction/ gram. exposure:				
M	37.15	27.85	14.23	7.92
SD	4.93	5.74	1.03	1.04
p	0.91	0.65	0.43	.045*
Round Two				
Gram. instruction/ vocab. exposure:				
M	39.29	28.43	14.57	8.29
SD	2.22	3.82	0.79	0.77
Vocab. instruction/ gram. exposure:				
M	34.86	27.14	13.14	7.71
SD	4.26	5.05	1.57	1.11
p	.037*	0.6	0.06	0.28

Note: maximum scores possible — grammar proficiency, 50; vocabulary knowledge, 52; reading comprehension, 16.

Both treatment groups across both rounds of data collection were found to be largely similar in terms of abilities and background, as shown in Table 4.1. In Round One of the data collection, the two treatment groups, grammar instruction and vocabulary instruction, were shown to be equivalent based on tests of grammar, vocabulary, and reading comprehension, though concerning years studying English, the vocabulary instruction group reported having studied English for a statistically significant longer amount of time ($t = 2.11$, $p = .045$). The grammar and reading comprehension scores were obtained from the grammar and reading exams, which have been discussed already. Vocabulary scores were obtained from a multiple choice test of 52 words selected by the experimenter from intermediate and upper level EFL text books in the *Headway* series and from a list of university level academic vocabulary (Nation, 1990). This vocabulary test also served to pre-test possible target vocabulary for inclusion in the study. Thus 12 of the 52 words from the pre-test were eventually used as target vocabulary. The vocabulary pre-tests are discussed more fully below in Section 4.5.3. For Round Two, a comparison between groups shows that the groups are equivalent concerning each of the factors shown in Table 4.1 except grammar proficiency, for which the scores for the grammar instruction group were significantly higher ($t = 2.44$, $p = .037$).

Table 4.2. Results of pre-tests for knowledge of the target grammar and vocabulary items and results for noticing ability for non-instructed items for both experimental groups.

Groups	Pre-test knowledge		Noticing ability	
	Grammar pre-test	Vocabulary pre-test	Grammar	Vocabulary
Round One				
Gram. instruction/ vocab. exposure:				
M	12.82	10.26	18.8	42.31
SD	16.88	11.36	8.35	22.17
Vocab. instruction/ gram. exposure:				
M	14.1	5.13	19.66	38.46
SD	14.98	8.01	11.25	20.56
p	0.84	0.2	0.82	.65
Round Two				
Gram. instruction/ vocab. exposure:				
M	4.76	1.19	9.52	28.57
SD	8.13	3.15	16.27	27.58
Vocab. instruction/ gram. exposure:				
M	0	3.17	3.17	22.62
SD	0	4.07	8.4	19.67
p	0.17	0.39	0.38	0.65

Note: pre-test knowledge scores are given in percentage correct, while noticing scores are presented in accuracy scores.

Furthermore, all groups were found to be equivalent on pre-test knowledge of target grammar and vocabulary items, as well as on ability to notice grammar and vocabulary as shown in Table 4.2. Scores for "noticing ability" were participants' scores for the "non-instructed" group items used in the noticing tests during the data collection. (This "non-instructed" group of items will be explained more fully in Section 4.5.1.1.) Scores are averaged across all three testing times. The non-instructed group items were neutral items which appeared in the reading texts used for the noticing test, and thus they provide a post-hoc way of measuring the general noticing ability of subjects on grammar and vocabulary.

4.2. TARGET GRAMMAR AND VOCABULARY

For each round of the study, six grammatical structures and 12 vocabulary items were selected for instruction and exposure. The choice of items was based on the following five criteria. First, in order to reduce any effects of previous instruction on noticing and learning, it was important that items be ones which students had not specifically been taught before. As not all students had had uniform instruction previous to entering the high school, and since students tend to engage in rather intense personal study of English and also study with private tutors, the possibility of previous encounter with the items could not be ruled out, but could only be reduced. This was done by previewing lists of potential grammatical structures and vocabulary items with the individual teachers of the students and the head English teacher for the year, as well as by checking through the most probable textbooks used by students in classroom study previous to entering high school and then eliminating those items which were likely to have been taught. Also, a teacher trainer at the English department of the local university with 10 years' experience teaching high school students was asked for an opinion concerning the possibility of previous instruction on potential target items. Since the participants in the study had undergone a rather thorough review of grammar in the course of the preparations for the entrance exam to the high school as well as in their first half a year in the school, the target grammar points chosen for this study were rather minor points which were unlikely to have been touched by the previous years of study. While students were indeed at an intermediate level and had not mastered all of the standard grammar points presented in textbooks appropriate for their level, they had received instruction in some form or another on most all major points covered by a general descriptive grammar, thus resulting in the need for the use of more obscure points for this study.

Second, beyond lack of previous instruction, items were needed for which participants could demonstrate little previous knowledge as shown on tests of grammar and vocabulary. Although it is unclear exactly what impact previous knowledge would have on noticing, it would clearly affect learning scores. For this reason all potential target items were pre-tested, and items showing the lowest level of knowledge were then incorporated in the study.

Third, it was necessary that items be at the appropriate level for participants so that explicit instruction could potentially be successful. Therefore, grammatical structures needed to be chosen for which students possessed the grammatical concepts and, if necessary, termin-

ology, to make the structures appropriate for instruction. The same was necessary for vocabulary, though in this case care was taken to choose words for which participants possessed either the requisite factual or conceptual knowledge to understand them, or had the knowledge of the Hungarian translation equivalent. To verify that the target items were at the appropriate level for the participants, the list of potential items was given to teachers, who were asked to comment on their appropriateness.

Fourth, it was necessary that target items be sufficiently different from Hungarian in order to reduce the possibility of transfer having an effect on learning scores. Again, although it is unclear what the effect of L1 similarity would be on noticing, the possibility for interference on learning scores is clear. Native speakers of Hungarian, a current Hungarian grammar (Kenesei, Vágó, and Fenyvesi, 1998) and the most recognized English-Hungarian bilingual dictionary (Országh, 1990) were consulted to verify that target items were sufficiently different from Hungarian.

Finally, with only general ideas derivable from the literature concerning what makes one structure or word more salient and noticeable than another, an attempt was made to include a relatively wide variety of items in the target grammar and vocabulary. For this reason, the target grammar contains large-scale clause-level ordering, complement subcategorization, and local word order. For vocabulary, words representing a wide range of parts of speech were sought using a list of relatively rare words (Nation, 1990) that participants were unlikely to have had contact with.

The resulting list of potential target items was then given to participants to pre-test their knowledge; the items showing the lowest level of knowledge were then used in the study. The 12 grammatical structures and 24 vocabulary items which were ultimately used (half in each round of data collection) are described in the following sections. By following the above principles, and considering the relatively high level of knowledge of English and large amount of exposure to the language that participants had had, most traditional grammar points and basic vocabulary had to be ruled out and, thus, the items selected through this process represent a somewhat unusual set, which will be described below.

4.2.1. GRAMMAR

The target grammar for each round is presented in Table 4.3 and can be broken down into the five general categories presented in the following sections.

Table 4.3. Target grammar items.

	Round 1	Round 2
Cleft and pseudo-cleft:	– cleft sentence emphasizing subjects – pseudo-cleft emphasizing verbs	– pseudo-cleft emphasizing the subject of *be*
Complements:	– possessive with gerund complement	– judgement verb + *"to be"*
Word dependent grammar:	– "*have* a verb" construction – "*give* a verb" construction	– *despite* + gerund – *recommend* + gerund
Causatives :	– *get* causative	—
Inversion:	—	– inversion with conditional – inversion with adverb
Note: blank cells, marked with a dash, indicate that the grammar point was not used in that round.		

4.2.1.1. Clefts and pseudo-clefts (3 items)

Clefts and pseudo-clefts generally involve the novel ordering of clauses coupled with additional grammatical changes so as to give emphasis or focus to particular information. Cleft sentences can bring into focus subjects, objects, or adverbials and are traditionally defined as having the form of the empty subject holder *it*, followed by the verb *be* with the clause containing the information receiving focus as its complement, followed by a relative clause (Quirk, Greenbaum, Leech, and Svartvik, 1985:1384-1386). Only clefts emphasizing subjects were included in the target grammar, an example of which is shown in example 1, below.

(1) *It was a book published in 1898 which had the greatest impact.*

Pseudo-cleft sentences are typically defined as having a nominal relative clause as subject followed by *be* and the superordinate clause (Quirk et al., 1985:1387). Included in the target grammar were pseudo-clefts which allow for focus to fall on the predicate (something which cleft sentences cannot do), as can be seen in example 2.

(2) *What he did was answer the phone that was ringing in a phone booth.*

Pseudo-clefts also differ from cleft sentences in that focus is placed on the superordinate clause rather than the subordinate clause.

Also included in the target grammar were pseudo-cleft sentences which allowed for focus to be placed on the subject of *be*, as seen in example 3.

(3) *What the problem is is that many people don't have telephones.*

Cleft sentences and pseudo-clefts were used as target grammar since they involve larger, clause-level movement and structuring of sentences, in contrast with other target structures described below. The particular pseudo-cleft structure in example 3 has previously been noted as a structure particularly difficult for students to learn (Robinson, 1995a) and yet, at the same time, it would appear to be a particularly salient structure for noticing with the highly unusual feature of *be* appearing twice, consecutively, and for this reason it was used as a target structure.

It is important to note here, and elsewhere in this section, that transfer can be ruled out in assisting participants in learning these structures. Neither cleft sentences nor pseudo-clefts exist as such in Hungarian. Emphasis in Hungarian is produced almost exclusively through the movement of the emphasized constituent to pre-verbal position with a verbal prefix (if required) appearing after the verb. Verbs are emphasized through maintaining verbal prefixes as prefixes or through moving the verb to the initial position in the clause. Examples 4, 5, and 6 show sentences emphasizing a subject, predicate, and the subject of a copula, respectively; they are the Hungarian equivalents of the cleft and pseudo-clefts from examples 1, 2, and 3.

(4) *Egy 1898-ban kiadott könyvnek volt a legnagyobb hatása*[1].
a 1898-in published book.DAT was the greatest impact.3sgPOSS
'It was a book published in 1898 which had the greatest impact.'

(5) *Felvette a telefonfülkében csengő telefont.*
up.picked.3sg the telephone.booth.in ringing telephone.ACC
'What s/he did was answer the phone that was ringing in a phone booth.'

(6) *Az a baj, hogy sokaknak nincs telefonjuk.*
that the problem, that many.DAT is.no telephone.3plPOSS
'What the problem is is that many people don't have telephones.'

4.2.1.2. Complements for particular categories of verbs (2 items)

The first item, example 7, is a gerund complement consisting of a genitive subject followed by a gerund. This complement can only follow certain verbs.

(7) *She remembered Jason's writing his number on the card.*

The possessive form used in this complement is optional and is virtually synonymous with the use of the non-possessed form, although there are cases, such as with the use of a pronoun, where the possessive form is more likely found (Quirk et al., 1985:1063). The construction with the possessive is rarely taught in intermediate level EFL texts, though the construction without the possessive and the issue of gerund or infinitive complements are generally well covered.

---

[1] The following abbreviations are used in the glosses of example sentences in this study: 2sg = second person singular, 3sg = third person singular, 3pl = third person singular, ACC = accusative case, CAUS = causative, DAT = dative case, INF = infinitive, and POSS = personal possessive suffix.

Again, transfer from Hungarian to English is not possible here as there is no possible Hungarian form using a possessive equivalent to example 7. In Hungarian a relative clause is used, which is also a possible option in English, as seen in 8 below.

(8) *Emlékezett arra, ahogy Jason leírta a számát.*
remembered.3sg about how Jason wrote the number.3sgPOSS.ACC
a kártyára.
the card.onto
'S/he remembered that Jason wrote his number on the card.'

The second items in this category are so-called "judgement *to*" complements, and consist of a verb in the main clause, the meaning of which suggests some kind of judgement, followed by a complement containing an infinitive form of a verb as predicate (Dixon, 1991:222-223). Although this construction accepts any kind of infinitive in the subordinate clause, only complements containing a *be* infinitive, as shown in example 9, were taught and presented in this study.

(9) *John considers Mary to be a good singer.*

This was done since *be* is most commonly the verb of the subordinate clause in this construction (p. 223), as well as to make the presentation of the grammar point more manageable in the limited amount of time available for instruction. Furthermore, the sense of the main clause verbs as presenting a judgement seems clearer by limiting the subordinate clause verb to *be*. Finally, in this construction *be* is optional in some cases depending on the particular semantic class of verbs used in the main clause, providing little or no meaning change (p. 232).

There is no class of verbs in Hungarian which allows the use of the infinitive of *be* in a complement accompanying a sense of judgement. In the particular case of *consider*, as is seen in example 10, the indirect object *singer* is found in the dative case with no accompanying verb *be*.

(10) *János Máriát jó énekesnek tartja.*
John Mary.ACC good singer.DAT considers
'John considers Mary to be a good singer.'

Both of the constructions in 7 and 9 were included in the target grammar as examples of subtle grammatical differences which could be easily overlooked or not noticed due to the small changes in meaning or style which they create. Although these constructions require particular classes of verbs in the main clause, no attempt was made to teach participants the details of the types of verbs which are allowed. The constructions were shown with a variety of verbs, and it was noted for participants that not all verbs could be used. On the achievement tests, verbs were used which participants had seen in the exercises.

4.2.1.3. Word dependent grammar (4 items)

This category involves words with specific grammatical elements associated with them. Four target constructions were used involving specific verbs and their potential complements.

The first two constructions are related: the so-called "*have* a verb" and "*give* a verb" constructions (Dixon, 1991:61), seen in examples 11 and 12, respectively.

(11) They had a talk about the situation.
(12) He gave the cord a pull.

The "*have* a verb" construction requires the use of a deverbal noun in the complement formed from an intransitive verb, while *give* requires an object plus a deverbalized noun formed from a transitive verb in the complement. Deverbal nouns used in these particular constructions are also known as "eventive objects", and their use signifies an informal style. (Quirk et al., 1985:750). Both the *have* and *give* constructions furthermore imply that an action was done for a short time or perhaps on a whim (Dixon, 1991:348, 341). Although participants in this present study may have known specific instances of the use of these constructions, they did not know the productive rule behind their use.

In Hungarian, there are no constructions which use eventive objects with general verbs such as *have* or *give*, nor is there any specific aspect used in Hungarian which would indicate something happening for a short time or on a whim.

The third construction in this category was the use of a gerund, as opposed to an infinitive, following *recommend*, as shown in example13a.

(13a) Doctors recommend drinking a lot of water.
(13b) Doctors recommend patients to drink a lot of water.

In this construction, if an indirect object is present, as in 13b, a following verb must be an infinitive, though if no indirect object is present, a gerund is required. This is in direct contrast with Hungarian, example 14, where the same construction requires the use of a relative clause containing the infinitive following the equivalent of a modal *need*.

(14) *Az orvosok azt javasolják, hogy sok vizet*
the doctors that.ACC recommend that much water.ACC
kell inni.
need drink.INF
'Doctors recommend drinking a lot of water.'

The fourth construction, shown in example (15), is the use of an NP containing a gerund as the object of the preposition *despite*.

(15) Despite being warned, he swam there anyway.

This construction was specifically taught with a focus on an action in the subordinated concession clause, thus requiring a gerund. This contrasts with Hungarian, example 16, where an inflected verb rather than a gerund is required in the relative clause.

(16) *Ott úszott annak ellenére, hogy figyelmeztették, hogy*
there swam.2sg that.DAT despite, that warned.3pl, that
ne ússzon ott.
don't swim.3sg there
'Despite being warned, he swam there anyway.'

These four constructions were included to test the noticing of grammar associated with particular lexical items, in this case, novel combinations of known grammar and words.

4.2.1.4. Causatives (1 item)

The causative use of *get*, example 17, was used as a target grammar item as an alternative to other verbs associated with the causative, such as *make*, that participants were likely to already know.

(17) He got his car repaired.

Again in this case, participants had probably encountered this particular use of *get* in specific cases, such as *get one's hair cut,* but were not familiar with its more productive use. Hungarian contrasts with English in that there is no periphrastic causative, only a quite productive morphological causative, as shown in detail in example18.

(18) *Megjavíttatta az autóját*
fixed.CAUS.3sg the car.3sgPOSS.ACC
'He got his car repaired.'

4.2.1.5. Inversion (2 items)

Two cases of subject-verb inversion were included in the target grammar. Example 19 shows subject-verb inversion used in the case of subordinate conditional clauses. Examples 20a and 20b show subject-verb inversion with a fronted negative element: a negative adverbial and a fronted negative object phrase, respectively.

(19) Should you want a ride, just give me a call.
(20a) Rarely had I seen such a film.
(20b) Not a single dollar had he spent.

These points were chosen as target grammar due to the fact that errors with inversion tend to be rather persistent in learners at the intermediate stage and beyond (Celce-Murcia & Larsen-Freeman, 1983:204), and thus may be considered items which are difficult to learn and

perhaps difficult to notice as presence or absence of inversion does not change the meaning in these cases. Inversion as such does not occur in Hungarian syntax, though due to flexible word order, subject and verb may occur in either order due to reasons of emphasis or focus, as seen in examples 4-6.

4.2.2. VOCABULARY

The goal for target vocabulary was, through pre-testing, to arrive at a list of polysyllabic words of similar conceptual and structural complexity, representing nouns, verbs, adjectives and adverbs which participants demonstrated little or no knowledge of.

Concerning the complexity of the words, it was considered important to use only abstract words for target vocabulary. This was done because a mix of abstract and concrete words would seem to give advantage to the learning and perhaps noticing of concrete words, which would likely be more easily learned with mnemonic strategies such as using visual imagery. Also, these abstract words were appropriate for the learners academic and language level. Similar arguments can be made for using polysyllabic words: they are more appropriate for the level of the participants, and are less likely to be remembered through the use of mnemonic strategies based on sound. Finally, concerning word classes, without preconceived notions of whether word class would affect noticing, words representing the four major classes were sought.

Table 4.4. Target vocabulary items.

	Round 1	Round 2
Nouns:	*propensity* *principle* *succession*	*bulk* *consent* *flurry* *incentive*
Verbs:	*betray* *deteriorate* *expose to* *sustain* *condense* *depict* *trigger*	*compel* *contemplate* *evoke* *enhance*
Adjectives:	*profound*	*ambiguous* *tangible* *cumbersome*
Adverbs:	*deliberately*	*intrinsically*

The resulting list of words, Table 4.4, shows relatively even conceptual complexity, though structurally there were exceptions with the monosyllabic *bulk* in Round Two, and the verb plus preposition *expose to* in Round One. Concerning word classes, words from each major class were represented, though they were not evenly distributed within and across each

round. These deficiencies were due to difficulties in finding a large enough pool of unknown words, as demonstrated on pre-tests, from which to select the target vocabulary. This will be explained further in section 4.5.3 when the vocabulary pre-tests are more fully discussed. Finally, it should be noted that none of the words have cognates in Hungarian, though each word has a translation equivalent.

4.3. DESIGN AND OUTLINE OF PROCEDURE

This study was based on a quasi-experimental design, using participants in four existing classes which were randomly assigned to be part of two treatment groups. Each treatment group received instruction in one area, grammar or vocabulary, and then also, at the same time served as the exposure group for the opposite area. That is, the grammar instruction group simultaneously served as the vocabulary exposure group, while the vocabulary instruction group served as the grammar exposure group. This was achieved, as will be explained below, through the use of identical reading texts (differing only in the highlighting of target items) for both groups during instruction. The same four classes of students were used for participants in Round One and Round Two of the data collection, with classes being assigned to the opposite treatments for Round Two than they were assigned in Round One.

The two separate rounds of data collection followed the same procedures, though with differences in target grammar and vocabulary. There were some differences between the two rounds of data collection in group composition, as described above, and also some differences in timing. An overall schedule of the timing of instruction and testing for the treatment groups, broken down by groups of existing classes, is shown in Table 4.5. Following pre-testing, treatment consisted of two class sessions in which participants received instruction in or exposure to target grammar and vocabulary. This was followed by three sessions of post-tests for both the noticing and learning of target grammar and vocabulary: an immediate post-test session occurring within three to four days of the last instruction session, and two delayed post-test sessions. Retrospections concerning participants' answers on the noticing test were done immediately following the first noticing tests for each round.

Table 4.5. Schedule for instruction and testing for both rounds of data collection.

Class	Instruction		Testing		
	Session 1	Session 2	Post-test 1	Post-test 2	Post-test 3
Round One					
Grammar 1	Day 1	Day 2	Day 6	Day 13	Day 43
Grammar 2	Day 1	Day 2	Day 6	Day 15	Day 44
Vocabulary 1	Day 1	Day 3	Day 6	Day 22	Day 51
Vocabulary 2	Day 1	Day 2	Day 6	Day 14	Day 44
Round Two					
Grammar	Day 1	Day 2	Day 5	Day 12	Day 23
Vocabulary	Day 1	Day 2	Day 6	Day 13	Day 23

The timing for the instruction and immediate post-test was almost identical for each group and across both rounds of data collection: the two instruction sessions occurred either on consecutive days or with one intervening day, and the immediate post-test occurred either on the fifth or sixth day, that is, between three and four days after the end of instruction. As can be seen, there was greater variation in the timing of the delayed post-tests though. For vocabulary group one in Round One, post-tests 2 and 3 occurred on days 22 and 51, rather than on days 12 or 13 and 43 or 44, as happened with the other groups. In Round Two, while post-test 2 occurred at a similar time to Round One, post-test 3 occurred on day 23 rather than on 43 to 51 as in Round One. Discrepancies in timing were due to scheduling difficulties encountered when arranging testing sessions.

Instruction and testing sessions took place during regularly scheduled class times and were conducted by the experimenter, with two assistants helping in the collection of retrospection data. The experimenter was not the regular classroom teacher of the participants, and to maintain student motivation, the regular teachers announced to students that they would be held accountable for material taught in the special sessions just as they would for material taught in regular class sessions.

4.4. INSTRUCTION AND EXPOSURE

Instruction and exposure for each round of data collection was carried out in two 45-minute class sessions. All reading and exercises were done in the class sessions themselves, with no outside-of-class work assigned to students. Furthermore, all materials (see Appendix A for an example of the teaching materials used) were in the form of xeroxed handouts and were collected from participants at the end of each session. This was done to remove the possibility that participants would study the materials at home, and also to let them know that their in-class written work done on the actual handouts themselves would be checked by the examiner.

All materials in the class sessions were presented in English, as was typical for classes in this particular high school. Hungarian was used in some cases, though: written instructions for exercises were presented both in English and Hungarian, and, occasionally, unknown English vocabulary was explained with a Hungarian translation equivalent.

Although there were two experimental class sessions for each round of data collection, difficulties with participants' attendance due to sickness or scheduling problems resulted in some participants only being able to attend one instructional session. To handle this potential problem, in each session all target items were presented or taught (depending on the group), and thus if participants missed one session, they were still instructed or received exposure in all target items. In Round One, one participant missed one instructional session in the grammar instruction group, while five participants missed one session in the vocabulary instruction group; in Round Two, all participants in both groups attended both instructional sessions.

4.4.1. READING TEXTS AND READING TASKS

As mentioned in section 4.3, instruction and exposure were carried out simultaneously in that each group which was being instructed in grammar was also being exposed to the target vocabulary through reading passages used in each class session, just as groups instructed in vocabulary were being exposed to the target grammar. This was done through the use of reading texts for each group which contained all target grammar and vocabulary and which were identical, save for typographical alterations, which will be discussed in the following section.

The texts for use as the foundation for the instructional materials were selected from intermediate and upper-intermediate textbooks which the participants were unlikely to have seen before. Criteria for choosing the texts were appropriateness of level, ability to be adapted for use with target grammar and vocabulary, length, and the possibility of the topic generating interest from the participants. Furthermore, well-structured texts were chosen which contained a plot or developed an argument, rather than simply presented information, so as to help maintain student attention and interest. The regular classroom teachers were consulted for advice in selecting the texts. Two texts were used for instruction in each round of data collection, with different texts used in each round. The texts used in the first instruction session were approximately 700 words in length, while the texts used in the second sessions were approximately 425 words long. These differences were due to time considerations during the class sessions.

The texts were then altered, edited, and sometimes rewritten, so as to include all of the target grammar and vocabulary items. Target items appeared only once in each text, and care was taken to attempt to place target grammar and vocabulary in the text in such a way that the comprehension of the key elements of the text did not depend on the target items. This was done to try to limit factors in the text which would make one item more salient than another.

The xeroxed pages of materials created for students began with the reading text, followed by five to seven comprehension questions (depending on the text) which focused on information found throughout the text. The questions were designed to make sure that participants had read the entire text and to give them an opportunity to check their understanding of the passage. Since students in both the grammar and vocabulary groups received the same reading text and comprehension questions, care was taken not to draw participants' attention to particular grammar or vocabulary via the comprehension questions. Following the comprehension questions, three discussion questions related to the text were included (though on one reading text these discussion questions were unintentionally omitted). Exercises for grammar or vocabulary then followed the reading, as will be described in the following sections.

The procedure for using the reading passages was the same for each group in both rounds of data collection. After a pre-reading exercise in which the topic of the reading was introduced, participants were instructed to read the passage silently and answer the comprehension questions by themselves. They were also told that the highlighted grammar or vocabulary in the text was important and that the class would be discussing and learning those items during the class session. After all participants had completed the reading and the comprehension questions, the text and questions were discussed in a whole class format. Individual partici-

pants were asked to summarize for the whole class sections of the text (a technique traditionally and frequently used in Hungarian foreign language classrooms). Answers for the comprehension questions were elicited and discussed, and particular issues which came up concerning the topic itself would be briefly discussed. The purpose of approaching the reading in this manner was to provide several opportunities for participants to focus their attention on the text.

It was anticipated that participants might ask questions at this point about unknown grammar and vocabulary found in the texts and that these questions might pose a problem if they focused an individual's or a group's attention on target grammar or vocabulary which they were merely to be exposed to. In practice this proved to be a rare occurrence. No questions concerning unknown grammar were asked by the vocabulary instruction group in either round, and most all questions concerning vocabulary asked by the grammar instruction group concerned key words needed for understanding the text, and not target vocabulary. This is most likely due to the fact that target grammar and vocabulary were placed in texts in such a way that they were not vital for comprehending the passage, and also due to the orientation that participants had by being told that their class would be focusing on either grammar or vocabulary. In this way then, lessons can be considered to have been exclusively devoted to either grammar or vocabulary, with little attention paid to the opposite area.

Table 4.6. Distribution of time in instructional sessions.

Grammar instruction			
Session 1		Session 2	
Approx. time	Activity	Approx. time	Activity
5	Introduction	5	Introduction
5	Pre-reading	15	Grammar exercise 3
10	Reading	5	Pre-reading
5	Post-reading	5	Reading
10	Grammar exercise 1	5	Post-reading
10	Grammar exercise 2	10	Grammar exercise 4
Total: 45 min.		Total: 45 min.	
Vocabulary instruction			
Session 1		Session 2	
Approx. time	Activity	Approx. time	Activity
5	Introduction	5	Introduction
5	Pre-reading	15	Finish vocabulary
10	Reading	5	exercise 2
5	Post-reading	5	Pre-reading
15	Vocabulary exercise 1	5	Reading
5	Begin vocabulary	10	Post-reading
Total: 45 min.	exercise 2	Total: 45 min.	Vocabulary exercise 3

Up to this point in each instructional session, the instruction and materials were the same for both the grammar and vocabulary groups, save for the differential highlighting of target items in the reading text depending on the focus of the group. After this, exercises designed specifically for each group began, as will be explained in the following sections. All exercises were contained on the handouts presented to students. The order of activities and the time spent on reading and exercises were the same for each group, though there was some slight variation in the pace across the different classes. An outline of the distribution of time spent during each session of instruction is presented in Table 4.6. In the second instruction session a shorter reading passage was used and fewer exercises were presented to allow time for completing the exercises from the previous class session.

4.4.2. GRAMMAR AND VOCABULARY INSTRUCTION

Each instruction session was centered around a reading passage in which the target grammar or vocabulary was highlighted through underlining, in the case of grammar, or using bold typeface, in the case of vocabulary. This text enhancement was done for two reasons. First it was done to allow for easy identification and discussion of target items in the context of the text, as is often done in foreign language instructional materials (e.g. Fuchs, Westheimer, and Bonner, 1994). Second, it was done in order to draw participants' attention to the target items as the texts were being read, and in this way potentially enhance the learning of the items. Text enhancement has been suggested as a technique to be used in grammar consciousness raising and has also been used in experimental studies where some positive effects on learning have been suggested (Alanen, 1995; White, 1998). Thus, the enhanced texts can be considered to be the first instructional input of the lessons. Note that only grammar was highlighted for the grammar group, and only vocabulary for the vocabulary group.

Following the reading, exercises were done which were specifically designed for grammar or vocabulary. Common to both grammar and vocabulary exercises was the intention to focus learners' attention on both the form and the meaning of the target items, and to present those items in the context of the reading passage when possible. The goal of instruction was to explicitly develop participants' declarative knowledge about the target items, first through identifying and understanding the target items, and then through quite controlled manipulation and production. No attempt was made to develop automaticity or fluency of use, or to practice the structures in communicative situations.

Each exercise was a written exercise presented to the participants on handouts. Depending on the exercise, participants worked alone or in pairs. Each exercise was introduced by the experimenter, and student work was monitored and checked during its completion. Answers were elicited and discussed after each exercise and participants were instructed to correct their own work while answers were being discussed. The pace of the exercises was monitored in such a way as to give participants the opportunity to think about the exercise and write down their answers before being asked to respond orally. This was done to avoid speeded drill-like exercises which might not give participants the opportunity to carefully and deliberately manipulate the knowledge needed for their answers. Such a context is suggested to be necessary for the establishment and practicing of declarative knowledge in the early stages of learning a particular item (DeKeyser, 1998).

4.4.2.1. Grammar exercises

Four different types of grammar exercises were conducted over the two days of instruction in each round (see Appendix B for an example). Exercise One and Two were completed on the first day of instruction following the first reading passage. On the second day of instruction, Exercise Three was completed, followed by the second reading passage and Exercise Four.

The first grammar exercise focused participants' attention on the form of the target grammar and also provided the opportunity for the experimenter to present and discuss the form of the items. This exercise consisted of 12 sentences, six of which had the same structure as the six target items in the text, and six which had similar but different structures. Participants worked individually and were instructed to match each target item in the text with one of the sentences from the exercise which had the same structure. As correct answers were elicited and discussed, the experimenter isolated the form of the construction on the blackboard and also discussed it using metalinguistic terminology which students were likely familiar with from previous discussions of grammar with their regular teacher.

Following this, the second exercise emphasized the meaning of the target constructions. The six target constructions and their line number locations from the text were presented once again in this exercise. Participants were asked to go back to the text, find the target construction, try to figure out the meaning of the construction from context, and then write a new sentence which means roughly the same thing as the target construction. These new sentences were to be written below each target item listed in the exercise. Students worked individually on the exercise. After the exercise was completed, correct answers were elicited and discussed, and participants were asked if they could see any differences in meaning between the target constructions and the roughly synonymous sentences which they had just written. This gave the experimenter the opportunity to present the rather subtle shades of meaning expressed by in the target constructions.

The third exercise was a sentence transformation task requiring participants to complete a sentence fragment using one of the target constructions in such as way that it has the same meaning as a sentence printed directly above it. This is a traditional exercise used in many EFL text books and is also used in the Cambridge language exams, and students were familiar with this type of exercise from these contexts. This gave participants the opportunity to produce the target constructions. The exercise was done individually, and after it was completed, correct answers were elicited and discussed by the experimenter.

The fourth exercise was another production exercise in which participants needed to create two original sentences using each target construction. In the exercise, participants were given a description of the construction and referred back to the reading text to see an example of it before writing their sentences. This exercise was done in pairs, and due to time constraints in all classes, pairs of participants were assigned to complete the exercise for only two of the six target constructions, though sentences for all six constructions were produced in the group as a whole. After the pairs completed their sentences, the experimenter elicited and discussed example sentences from all six target constructions.

4.4.2.2. Vocabulary exercises

Three vocabulary exercises were carried out in the vocabulary instruction sessions. The goal of the instruction was to emphasize the form and meaning of the vocabulary and to give participants the opportunity to experience and manipulate the words in a variety of contexts. Following the first reading passage, Exercise One and most of Exercise Two were done. In the second session, Exercise Two was completed, and then, following the second reading passage, Exercise Three was completed. All exercises were done individually by students.

The first exercise required participants to find the meaning of the vocabulary words through the context of the reading passage. In the exercise, twelve definitions were presented preceded by blanks. Participants were instructed to find the highlighted words in the text which matched each definition and then to write the word in the space provided. An example exercise was done with the participants in order to show ways in which context could be used to determine the meaning of words. Participants worked individually, and after the exercise was completed, the experimenter elicited answers and discussed them.

The second exercise was a cloze task in which participants were asked to use the highlighted words from the text to fill in gaps in sentences. One sentence was provided for each target vocabulary word. Following the completion of the exercise, corrected answers were elicited and discussed, including cases where multiple correct answers were possible.

The final exercise required students to memorize the Hungarian equivalents of the target vocabulary items. In the exercise, a list was presented to participants with the target vocabulary on the left side and the corresponding Hungarian words on the right. Due to time constraints in each class at this point in the lesson, participants were given three minutes to memorize the list and then were tested by the experimenter saying the English word and eliciting the correct Hungarian word from the group. Participants were asked to keep track of how many items they had correctly recalled. This exercise was used to quickly introduce and establish the Hungarian counterparts for the target vocabulary and thus to reinforce the meaning of the words which had been discussed in the previous exercises. This was done as an alternative to glossing the text with the Hungarian words, which could have had the effect of increasing the possibility that the grammar and vocabulary instruction groups would read the texts differently. In the vocabulary instruction groups the Hungarian equivalents had been mentioned on occasion before as a guide in understanding the meanings of words, though Exercise Three was the first systematic pairing of the Hungarian and English words.

4.4.3. EXPOSURE TO GRAMMAR AND VOCABULARY

As mentioned previously, each group served simultaneously for both instruction and exposure. Participants were exposed to one domain, grammar or vocabulary, through reading and completing exercises which were designed to focus attention on the other domain. Thus, exposure can be defined as having occurred through the reading of two texts and through the completion of exercises which required reference to the texts. Attention was not explicitly directed to the items intended for exposure, but, in fact, was directed either at comprehension of the reading passages, or at the items intended to be the object of instruction.

4.5. INSTRUMENTATION

4.5.1. THE NOTICING TEST

The test developed and used in this study to measure noticing is based on the tests and procedures used in recognition memory research to elicit data concerning participants' states of awareness of memories based on the Remember/Know distinction introduced by Tulving (1985) and developed by Gardiner (1988), as was discussed in the literature review. In these tests, participants are trained to give a "Remember" response when they have a conscious memory of having seen a test item on the study list, and give a "Know" response when they simply have a strong feeling that a test item appeared on the study list but cannot consciously recall it being there. As discussed in the literature review, the usefulness of this distinction in SLA research is that a Remember response represents a conscious memory at the time of the encoding and thus may possibly be used as evidence of having consciously noticed a particular linguistic element. Thus, the goal for the noticing test used in this current study was to present participants with a context in which they could, when appropriate, consistently report Remember responses concerning grammar and vocabulary which they encountered in input — specifically in reading — since it is these Remember responses which are used as the measure of noticing.

Recognition memory tests of this kind typically have the format of a study phase, where a list of words is presented, and a testing phase where the study words are presented with a series of distractors. The length of time between study and test as well as conditions at the time of study and test can be manipulated for various purposes. The noticing test developed for this current study differs from traditional tests in several important respects. First, along with individual words, memory for grammar is also being tested, though expanding the test beyond the use of word lists is not without precedent, as studies using the Remember/Know distinction have been previously carried out, concerning for example, for musical melodies (Gardiner, Kaminska, Dixon, and Java 1996) and material learned in academic courses (Conway et al., 1997).

A second and more significant way in which the noticing test appears to differ from traditional tests using the Remember/Know paradigm is that what is being tested is not simply memory for items which were instructed, but the impact of instruction on the later encoding of memories of those instructed items when they are encountered in reading texts. The question being investigated is whether instructed items are somehow given familiarity and salience, thus making a later encounter with them, more likely to be accompanied by consciousness of these items. A design of this type is actually not without precedent either, as, for example, memories for low and high frequency words have been studied using the methodology described here (e.g. Dewhurst, Graham, and Barry, 1998) where the effect of words' frequency of previous encounter on participants' encoding at the time of study was measured. This is essentially what is being done here in the current study, except that the previous encounter with target items is being experimentally controlled through assigning groups to exposure or instruction conditions. For this reason, the noticing test basically consists of a short passage for participants to read, followed by a list of test items contained in the text along with dis-

tractors. Participants then mark the items in the list as having appeared in the text or not, and then make the further distinction of marking Remember or Know or Guess next to the marked items depending on whether their recollection of the items is accompanied by conscious memories or not. In this sense, then, the "study" phase of the test occurs at the time of reading the text used in the test for noticing, which would correspond, in the typical test design to the time when participants initially memorize the list of words. The "testing" phase, then, occurs when the list of previously encountered items and distractors is reviewed by participants.

4.5.1.1. Design of the noticing test

As was mentioned above, the basic design of the test consisted of a reading text containing target grammar and vocabulary (the study phase), and a subsequent list of items (the testing phase) containing target items and distractors on which participants indicated whether or not they believed the items appeared in the reading text which they had just seen. In this design the exact same testing materials were used for both experimental groups. Four important issues need to be discussed in relation to the design of the noticing test: 1) the creation of the reading texts used in the study phase of the test; 2) the items included in the text used in the study phase and the items and distractors placed on the list used in the testing phase; 3) the presentation and layout of the lists used in the testing phase; and 4) the possible responses which participants were allowed to make to items on these lists.

First, the choice of texts used for the study phase of the noticing test and the method in which the texts were altered to include target items followed the same principles used in the creation of the texts used for instruction, as described in section 4.4.1. (See Appendix C for an example of a text used in testing.) The texts were selected from the reading comprehension sections of two Cambridge First Certificate exam practice books (Hashemi, 1991; University of Cambridge Local Examinations Syndicate, 1990), which, as mentioned before, were presumed to be at the appropriate level for participants. The selection of texts was based on length, the ease with which texts could be altered to include target grammar and vocabulary, and the presumed accessibility of the topic for participants. Texts of approximately 500 words, amounting to around one double spaced page, were chosen so as to provide a text which would take participants from three to five minutes to read. The amount of time it took to read a text of this length was appropriate for the overall time limitations of the testing sessions, and also seemed to be an appropriate period of time for sustaining participants' attention to the task. Texts were chosen with topics that participants were assumed to have the required background knowledge for and which were well organized either around a narrative or the clear development of a central idea or series of ideas. Finally, all texts were presented with four true/false comprehension questions at the end of the text. These questions were altered versions of multiple-choice questions contained along with the texts in the exam books.[2] Questions were included so as to maintain participant attention and require them to read the entire text.

---

[2] Participants' answers to these comprehension questions formed the basis for the reading comprehension scores described in section 4.1.

Second, in order to test the hypotheses presented in Chapter 3 and feel reasonably confident that the data are valid, responses concerning four different categories of items needed to be collected and thus were present on the list of items used in the testing phase. The group termed "target items" was used to measure the key dependent variable of noticing, while three other groups of items were used as checks to make inferences about the validity of responses to the target items, as will be explained below.

Grammar and vocabulary which were the object of instruction and exposure were divided for testing purposes between two groups: "target items", which appeared on both the text used in the study phase of the noticing test and on the list in the testing phase, and "+target/-text" items, which did not appear on the text used in the study phase of the noticing test, but only appeared on the list used in the testing phase. Responses to the target items provided the data used to test the central hypotheses of the project, while +target/-text items allowed for the testing of the possibility that participants were merely reporting that they remembered those items which were the object of instruction and exposure, rather than items which actually appeared in the text in the study phase of the noticing test. In order to test for the possibility that participants were simply saying that they noticed items which they were instructed in but not in the text used for noticing, for each noticing test done a certain number of items which were the object of instruction or exposure needed to be set aside for the use as +target/-text items. Of the 12 vocabulary items used in each round, three were set aside for use as +target/-text items, thus leaving nine items in the target items category. Of the six grammar items which were the object of instruction or exposure in each round, one was set aside for use in the +target/–text item group, leaving five items for the target items group. In order to reduce the possibility of a testing effect in which participants might anticipate which items were in the target and +target/–text groups, different items were rotated into the +target/-text group for each post-test in each round. The result of this is that the exact composition of the target and +target/–text groups was different across the three post-tests of each data collection round.

A third group of items, labeled "non-instructed", were grammar and vocabulary which were not the object of instruction or exposure, but which appeared in the text used in the study phase of the noticing text and also on the testing list.[3] In order to select these items, the basic texts which were the foundation of the study phase texts were consulted, and items which appeared suitable for testing were selected for the non-instructed group items. The basic texts themselves, as was explained above, were taken from the reading tests found in First Certificate, or intermediate level, exams, and thus the grammar and vocabulary contained in them was considered appropriate for the proficiency level of the participants, and it was this grammar and vocabulary suggested by these tests which was used for the non-instructed group of items. As these items were not pre-tested, they cannot necessarily be considered to be equivalent to target items, which were pre-tested and thus represent a relatively unknown group of grammar and vocabulary. The reason for including the non-instructed items was to build evidence for the central hypothesis concerning instruction posit-

[3] It is important to note that the label "non-instructed" refers throughout to items which were neither the object of instruction *nor* exposure.

ively affecting the noticing of instructed items in that if, within the instructed group, scores are lower on non-instructed items than target items, it would in some part add support for the possibility that instruction made items more salient and more susceptible to noticing. Furthermore, a comparison between groups on non-instructed items would allow for the description of whether the two experimental groups were equivalent in their ability to notice grammar and vocabulary, as was done in Table 4.2.

The fourth group of items, "distractors" were neither the object of instruction nor exposure and did not appear in the text used in the study phase of the noticing test, but were used as the distractors which the target, and non-instructed items were paired with. Thus, participants' indication of having noticed items from the distractor group are errors, as this indicates that they did not notice the correct item. Distractors are different from +target/–text items, although the choice of either is an incorrect answer. The +target/–text items are grammar or vocabulary which participants were taught or exposed to and which were then paired with non-instructed items which did occur in the text for noticing. The purpose for this was explained above.

0 enhance	0 remember 0 know
0 improve	0 guess/not sure

0 Our guide cut a narrow path...	0 remember 0 know
0 What our guide did was to cut a narrow path...	0 guess/not sure

Figure 4.1. Examples of the layout of testing items on the noticing test for grammar and vocabulary.

Third, for the lists of items used in the testing phase of the noticing test, each item was presented in a two-alternative forced-choice format with one correct option which had appeared in the text used in the noticing test and one incorrect option which had not appeared there. On the page, each two-item pair was isolated in a box to allow them to be clearly seen and separated from other items. Participants were instructed to choose which items had appeared in the previous reading text, and then to mark whether this memory was accompanied by conscious recollection or not, or whether they were simply making a guess. Figure 4.1 presents an example of the layout for testing items. For vocabulary, correct choices were presented individually, paired with near-synonym incorrect choices. For grammar, phrases were presented which contained the target grammar, and these were paired with phrases expressing nearly the same meaning and using as close to the same vocabulary as possible. The order of the pairing of correct and incorrect choices was randomly presented.

The forced-choice format was used for two reasons. First, this appeared to be the best way to test the noticing of grammar in that it reduces the possibility that participants were merely remembering the lexical items used in the grammatical constructions rather than the grammar

itself. By presenting together correct choices and incorrect choices using almost identical lexical items, it is the grammar which contrasts between the two items. This focus on grammatical structure would have been much more difficult to ensure if items were presented in a list format. Second, the forced-choice procedure, a technique frequently employed in recognition memory tests, was chosen to reduce the possibility of response bias whereby participants' hesitancy or over-enthusiasm to report memories for items would influence their scores (Green & Swets, 1966:408). Using the forced-choice format, participants' overall scores (Remember, Know and Guess considered together) for selecting either correct choices or incorrect choices are much less likely to be influenced by a tendency to under- or over-report memories for particular items.

Fourth, as shown in Figure 4.1, three options were presented on the noticing test for indicating a memory for an item: Remember, Know, and Guess. To review, a Remember response represents a retrieval from episodic memory, while a Know response represents a retrieval from semantic memory (Tulving, 1985), and thus Remember responses are seen to represent cases of noticing, that is the conscious experience of a particular grammar or vocabulary item. The issues surrounding Remember and Know responses have been discussed above in this section and in the literature review. The possibility for a Guess response was included so that participants would not be forced to choose an option indicating a conscious memory or a strong feeling when neither was present. The advantage of including a possibility of a Guess response — beyond the obvious increase in the face validity of the procedure — is that when guessing is not an option, guesses tend to be included in Know responses (Gardiner, Richardson-Klavehn, and Ramponi, 1997: 391), thus reducing the validity of the response. In experimental settings Guess responses themselves have been shown to reflect little knowledge of studied material (Gardiner, Ramponi, and Richardson-Klavehn, 1998), though outside of experimental settings, for example concerning memories of academic material encountered in classroom settings, Guess responses do reflect some accuracy due to the use of strategies in evaluating incorrect choices based on the abundance of contextual information provided in an academic course (Conway et al., 1997).

In summary, then, the noticing test consisted of a study phase where participants read a short passage containing all target grammar and vocabulary, and a testing phase where, through a forced-choice procedure, participants indicated different types of memories, or lack of memories, for test items found in the reading by marking Remember, Know, or Guess depending on their experience of the item.

4.5.1.2. The procedure for the noticing test

During the testing sessions, the noticing test was given first, followed by the tests of learning. First, participants were given a set of written instructions for the noticing test (see Appendix B for a copy). These instructions were considered crucial in guiding participants to make accurate and appropriate responses. The instructions, presented both in Hungarian and English, described the format of the test, explained the difference between Remember and Know responses, and then described example situations where Remember and Know responses would be appropriate. Examples covered both the testing situation, where memories

for grammar and vocabulary were concerned, and everyday situations, which were presented to back up the Remember and Know distinction in general. This discussion of the instructions for the first testing session lasted between five and ten minutes, with shorter amounts of time being taken in later sessions.

After the instructions were read and discussed, the study text was handed out to participants, and approximately five minutes were allotted for the reading of the text. After all participants had completed the reading, the texts were collected and a three-to-five-minute unrelated filler activity was done so as to not allow participants to rehearse parts of the texts, thereby not allowing them to keep the text active in working memory. Following this, the handouts listing the grammar and vocabulary for the noticing test were distributed. Participants were given as much time as they needed to complete the test. All participants finished the test within 15 minutes, with some finishing it in as little as five minutes. After participants completed the test, they were instructed to check over their answers to make sure that they had filled in each answer and to make sure that they used the appropriate reasons for indicating Remember, Know or Guess. Following this, as will be explained in the next section, during the first test of each data collection round, retrospections were collected from participants concerning their reasons for their responses.

4.5.1.3. Steps taken to ensure accurate responses on the noticing test

Due to the fact that the noticing test depended on participants accurately reporting their subjective internal states of awareness of target grammar and vocabulary as encountered in the reading passage, and that the reporting of this kind of information without practice may be difficult (Cohen, 1987), steps were taken to train participants to use the noticing test accurately and to check their responses at the time of testing. This was done in three ways.

First, prior to the first noticing test, participants took part in a formal training session integrated into the instruction sessions during which the format of the exam was explained, the crucial difference between Remembering and Knowing was discussed, and a short sample test was taken.

The formal training for the noticing test took part across the two instructional sessions. At the end of the first instructional session participants were told that in a few days they were going to be taking a "new kind of test" and that they would need to understand how the test worked. The experimenter then briefly described the test and gave the participants a detailed set of instructions for the test, the reading of which was assigned for homework. These were the actual instructions for the noticing test which were also handed out and used during each testing session. At the end of the second session of instruction, a 15-minute training and practice session for the noticing test was held. During the session, the Remember and Know distinction was presented and discussed by the experimenter, and participants were given another chance to read the instructions for the test. After answering participants' questions, a short practice version of the noticing test was given consisting of a 175-word reading passage and 3 grammar items and 4 vocabulary items for testing. After the practice test was completed, the experimenter led a discussion of the answers, eliciting and evaluating participants' explanations for their Remember, Know and Guess responses.

Second, as was indicated in the previous section, before each noticing test, instructions for the noticing test were again handed out for participants to read, and the Remember and Know distinction was again briefly presented and discussed. Furthermore, participants were encouraged to refer back to the instructions at any time during the test or to ask questions of the experimenter concerning the test.

Third, during the first test of each round of data collection, verbal retrospections were collected from participants concerning their reasons for making Remember, Know or Guess responses. There were three reasons for collecting the retrospections. First, this was done to directly verify the validity of participants' responses and to make any necessary changes, as discussed below. Second, the retrospections gave participants the opportunity to discuss one-on-one the Remember/Know distinction and ask any questions that they might have. And third, it was hoped that the retrospections would encourage participants to take the test seriously through the motivation provided by directly checking their answers.

As participants finished the noticing test, the experimenter and two assistants talked with each participant individually and audiotaped their explanations for their answers. Recognizing the fact that participants may not be able to articulately express or describe an internal state of awareness (Ericsson & Simon, 1993:45), participants were asked a series of questions intended to help them identify what, if anything, they were conscious of when encountering specific grammar and vocabulary. When participants' explanations were not appropriate for the responses they made, the responses were discussed and changed. The experimenter's assistants were trained to take the retrospections by reviewing retrospections which had been collected during the pilot project for this research. During the review of the pilot project retrospections, a certain set of categories were found which were typical of Remember and Know responses. These categories were used for anticipating responses which would be encountered and for preparing the experimenter and assistants for the retrospections. (A discussion of the categories of responses will be presented in section 5.7.1.2.)

4.5.2. GRAMMAR TESTS

A sentence completion task was used in both grammar pre-tests and post-tests. This task was also used as an exercise in both grammar instruction rounds, as explained in section 4.4.2.1. The format of the task involves the presentation of an example sentence followed by a sentence stem which participants were required to complete so as to have nearly the same meaning as the example sentence. One sentence completion exercise was presented for each target grammar item, making a total of six sentences of target grammar items for participants to complete.

As a second test, a grammaticality judgement task was given for the post-tests only. Participants were presented with 12 sentences, one grammatical and one ungrammatical sentence for each target grammar item, and were instructed to mark which ones were correct and which incorrect.

Prior to testing, participants were given instructions concerning each exercise type, and during testing the experimenter answered participants' questions and spot-checked their work.

4.5.3. VOCABULARY TESTS

A word definition task was used for vocabulary post-tests. In this task a list of the 12 target words was presented with a blank beside each word. Following the testing procedure used by Watanabe (1997), participants were instructed to provide a Hungarian translation equivalent of the word in the blank provided, and if they were not able to do that, they were asked to provide any information they could about the word even if that was just a general indication of some aspect of the word's meaning. This procedure was used so as to potentially tap partial information about the word that participants possessed in absence of a translation equivalent or clear definition. (See section 4.6.3 for details concerning the scoring of this test.)

Concerning the pre-tests, there was an unintended mismatch of pre- and post-test types for Round One, while Round Two used consistent pre- and post-tests. The vocabulary pre-test used in Round One was a matching task where words were matched with their translation equivalents, while the pre-test for Round Two vocabulary was the definition task described above. This mismatch was an unintended consequence of using a too simple pre-test design in Round One. In this matching test four to six Hungarian words were to be matched with their English equivalents from a pool of seven to nine words. Following Nation (1990), it was assumed that a multiple-choice test using a larger number of distractors would reduce the possibility of guesses while providing a convenient way of testing a large number of words. Yet, due most likely to the way that words were grouped (combining both higher and lower frequency words together) and participants' great amount of exposure to English, participants were able to easily match definitions for almost all of the words, leaving only a few words which participants had not marked correctly. These 12 words were used in the first round as target vocabulary, and also at that time it was decided to abandon this method of testing vocabulary and adopt a test, the definitions test, which would place greater demands on participants to demonstrate their knowledge. The result was the mismatch of pre- and post-tests for vocabulary in Round One. It should be pointed out, though, that since the matching format pre-test most likely allowed for a greater possibility of guessing and inferencing than did the definitions test, that the level of actual knowledge reflected in the resulting group of unknown words produced by the matching pre-test was indeed small.

4.6. SCORING AND ANALYSIS

4.6.1. SCORING THE NOTICING TEST

In scoring the noticing test, accuracy scores were calculated separately for each response type, Remember, Know, and Guess. That is, the accuracy of each response type was calculated independently of the other responses.

Accuracy scores, essentially involving a correction for error, as will be explained below, were used for calculating all noticing scores throughout the study rather than using the simpler measure of the percentage correct. The reason for doing this was that participants, when encountering a pair of items on the list used for the noticing test, not only made a decision about which item they believed they had noticed, but had the three options for responding, Remember, Know or Guess. While the forced-choice procedure can minimize response bias in the selection of one item over another, it does not take into account the three possible ways of responding to an item. Thus, calculating the percentage correct, rather than an accuracy score, could be misleading, as the following example shows. Without correcting for error there could be a situation where two participants have the same scores for percentage correct for Remember responses, say 30%, yet their responses could have been qualitatively quite different. For one participant, this could mean 30% correct Remember responses with the other 70% of the responses (correct or incorrect) being distributed across the Know and Guess responses. For the other participant, a score of 30% correct could be achieved and the remaining 70% of the responses could have been *incorrect* Remember responses. The second participant is actually much less accurate in making Remember responses than the first, something which would not be captured by simply calculating the percentage correct. Thus, when a participant chooses a Know or Guess response rather than a Remember response, this is a qualitatively different decision than making an incorrect Remember response. What was needed was a measure which reflects the sensitivity of a participant's ability to recall items, and for this reason, steps were taken to create a score for each response, Remember, Know, and Guess, which would reflect the level of accuracy of judgments.

The technique used to adjust scores to reflect accuracy was to subtract instances of choosing incorrect items, that is, false alarms, from correct responses, that is, hits. This technique of obtaining an accuracy score by subtracting false alarms from hits has been employed in previous research using the Remember/Know paradigm (e.g. Gardiner et al., 1997; Strack & Förster, 1995) and has been used in research on recognition memory in general for several decades (Snodgrass & Corwin, 1988:38). Thus, scores were figured separately for Remember, Know or Guess responses by calculating a participant's overall correct answers for a particular response and then subtracting the overall number of incorrect answers for that response type. The result are scores for Remember, Know and Guess which reflect the accuracy for each of the responses.

4.6.2. SCORING THE GRAMMAR TESTS

For the sentence completion task, one point was given for each correct answer. Each sentence stem was set up so that one target structure could be used to complete it. Correct answers were those which used the target structure in a grammatically correct way. Grammatical errors in other parts of the sentence were ignored.

Adjustments in scores needed to be made to account for correct answers which did not incorporate the target grammar. Although care was taken to try to create sentence stems which could only be completed using a target structure, participants were still able, on occasion, to form grammatically correct sentences with similar meaning to the example sentence, but by using another construction. These responses are problematic in that they do not reflect knowledge of the target construction, though they do not point to lack of knowledge as a clearly incorrect or unfinished answer would. Thus, for each individual participant, when a correct sentence was created not using the target structure, that item was ignored and the total number of possible items for that participant was reduced by one. Three example responses are presented below in 21-23. Example 21 is a grammatical sentence not using the target form and yet acceptable, and thus was not counted into the calculation for total correct responses for this participant. 22 is an example of an ungrammatical, and thus incorrect, answer. And 23 is an example of a grammatically correct sentence which was marked incorrect because its meaning diverges greatly from the example sentence. (Note that participant responses are underlined.)

(21) *Mary recommends everyone to eat spinach.*
Mary recommends spinach to everyone.

(22) *Richard considers that Jane is qualified.*
Richard considers Jane qualification.

(23) *He admired that John reported the incident.*
He admired John's report of the incident.

Scores for the grammaticality judgement task were calculated by giving one point for each correct response. Scores for grammatical sentences and ungrammatical sentences were figured separately.

4.6.3. SCORING THE VOCABULARY TESTS

The multiple-choice vocabulary test used in the pre-test for Round One was scored giving one point for each correct answer.

The definitions test was scored following Watanabe (1997), who used a three-point scoring system to evaluate participants' answers. In this system, a correct definition or translation received three points, a partial definition received two points, and an answer which reflects some correct information about the word was given one point. A native speaker of Hungarian was used as the judge for the definitions test. 24-27 are examples of answers and how they

were scored. (Note that participant responses are in boldfaced type in these examples.) In 24, no points were given as the answer is a mistranslation of the word. 25 received one point for providing information about the sense of the word *succession* as having to do with things happening in a particular order. Two points were given for example 26 for providing a close but incomplete translation. And 27 received three points for exact translation equivalent of the target word.

(24) succession: ***siker*** "success" (0 points)

(25) succession: ***folyamat*** "process" (1 point)

(26) deteriorate: ***elront*** "ruin" (2 points)

(27) depict: ***ábrázol*** "depict" (3 points)

4.6.4. ADJUSTING GRAMMAR AND VOCABULARY SCORES FOR PREVIOUS KNOWLEDGE

Scores for learning were used for two purposes, to compare instruction and exposure groups and to look at possible relationships between noticing and learning; both instruction and exposure groups showed evidence of previous knowledge of target items in each round, yet this knowledge needs to be considered differently in each case, either comparing groups on learning scores or in calculating correlations between learning and noticing. Since, as was shown above in Table 4.2, previous knowledge was evenly distributed between groups, with no significant differences found, there is no need to factor out previous knowledge when making comparisons between groups on their learning of grammar and vocabulary. In plotting individual students' scores for learning and noticing though, it was necessary to make adjustments for previous knowledge since this previous knowledge could allow for the possibility of higher learning scores which were not accounted for by instruction or exposure during the instructional sessions. This knowledge, then, could unnaturally skew the relationship between noticing and learning. In order to adjust for previous knowledge of target items, items which participants showed previous knowledge of were removed from an individual's score, and the total possible score for that individual was reduced by one.

4.6.5. ANALYSIS OF THE RETROSPECTIONS

Although the main purpose of the retrospections was to improve the validity of participants' responses and thus was fulfilled at the time of data collection, a further analysis of the retrospections was done in order to describe participants' reasons behind the Remember, Know and Guess responses. Reasons and explanations were categorized according to the set of categories which was established through the pilot project for this study, and further categories were also added. Details of the categories can be found below in the discussion of the results in section 5.7.1.2. below.

CHAPTER 5.
RESULTS AND DISCUSSION FOR NOTICING

The presentation of the results and discussion will be organized in the same way in all chapters. Results for both data collection rounds and from the three different testing times within each round will be considered together within each section, with grammar and vocabulary being given separate sections. The results and discussion sections will be divided into four chapters: group data for noticing, group data for learning, noticing data for individual target items, and noticing and learning data for individual participants. In the first two chapters, data will be presented and discussed which relate directly to the hypotheses of this project, and in the final two chapters data will be presented which add support to the hypotheses and develop additional points. After the presentation of results, there will be a discussion in each of the four chapters. A general discussion will then follow.

In this chapter, and all following chapters of results and discussion, data will be presented from both rounds of data collection and from the three different testing times within each round. Due to the amount of data being presented, it is important at the outset to again note two things. First, in the two data collection rounds, different groups of participants were used as well as different target vocabulary and grammar. Second, within each round at each of the three testing times, a slightly different group of target grammar and vocabulary items were tested — something which was done to control for a testing effect, as was noted above. For these reasons, it will not be possible to make direct comparisons across the rounds of data collection, nor will it be possible to make direct comparisons across the testing times within each round. Therefore, the different data collection rounds and the tests done at different times within these rounds would best be thought of as replications which look at the same questions, but with different groups of participants (in the case of the two data collection rounds), slightly different groups of structures (within each round), and then over different amounts of time between testing and treatment. This situation should not preclude noting general similarities and trends, but will not allow for direct testing across times and between rounds.

Two points should also be made concerning the statistical analysis and the presentation of these results. First, it should be noted that concerning the interpretation of the statistical results, an alpha-level of .016 was set using the Bonferroni correction in order to control for the Type 1 error rate due to the multiple use of t-tests. In doing this it was necessary to define the domain in which the Type 1 error rate would be controlled, and this domain was defined as the testing of the same participants on the same variable across the three testing times. Thus, the traditional .05 alpha level was divided by three, making .016 the result. To make

reading the statistical results easier, all results significant at the .016 level will be marked with an asterisk. Unless otherwise noted, the alpha level of .016 will hold throughout the text. Second, in all cases unless otherwise indicated, one-tailed tests were used in this project. One-tailed tests were used due to the fact that all hypotheses used in this project were directional. Specific predictions were thus being made, and in this case one-tailed testing is appropriate (Hatch & Lazaraton, 1991:230-231).

Finally, it should be noted again that all of the scores presented for the noticing results are expressed in terms accuracy scores rather than as a percentage correct, as was explained in the methodology section.

5.1. OVERVIEW OF RESULTS CONCERNING NOTICING

Tables 5.1 and 5.2 present an overview of the results for noticing in Round One and Round Two for the target items group for grammar and vocabulary. These tables can first serve as a review of the types of data that were collected, and, second, they can provide an overview of the general trends that are found in the data, which will be discussed in more detail in other sections.

Table 5.1. Summary of Grammar noticing scores for the target items for Round One and Round Two. Scores for Remember, Know and Guess responses are provided.[1]

	T1			T2			T3		
	R	K	G	R	K	G	R	K	G
Round One									
Instructed group									
M	38.46	0	3.08	23.08	1.54	1.54	24.62	6.15	–7.69
SD	23.75	14.14	32.5	17.97	15.19	27.64	26.02	27.55	33.2
Exposure group									
M	10.77	–6.15	–18.46	20	1.54	–7.69	10.77	0	–18.46
SD	22.53	25.01	32.11	25.82	35.08	27.74	22.53	23.09	28.82
Round Two									
Instructed group									
M	42.86	25.71	–14.29	20	5.71	11.43	17.14	8.57	17.14
SD	26.9	15.12	19.02	16.33	27.6	36.25	17.99	15.74	31.47
Exposure group									
M	11.43	–2.86	–11.43	5.71	17.14	14.29	20	2.86	–2.86
SD	27.95	7.56	57.57	19.02	17.99	51.27	23.09	21.38	37.29

[1] Note that all noticing scores presented in the tables throughout this text are accuracy scores expressed in terms of a percentage ranging from 100% accurate (+100) to 100% inaccurate (-100) unless otherwise noted.

First, at each testing time, T1 though T3, data were collected on the three responses of Remember, Know, and Guess. Most important for this project, though, are the Remember responses. Although Know and Guess responses will be discussed in some sections, mainly Remember responses will be reported in future tables, unless otherwise noted. Second, it is important to note that the instructed group for one area, say vocabulary, is the exposure group for the other area, in this case, grammar. Thus it is possible to compare the same groups on different variables. And third, since accuracy scores are being calculated, negative scores are possible, which show that there were more incorrect than correct answers given for a particular group.

Table 5.2. Summary of Vocabulary noticing scores for the target items for Round One and Two. Scores for Remember, Know and Guess responses are provided

	T1			T2			T3		
	R	K	G	R	K	G	R	K	G
Round One									
Instructed group									
M	54.7	19.66	13.68	37.61	13.68	16.24	43.59	19.66	12.82
SD	28.5	22.75	13.72	21.53	19.33	17.92	26.24	12.13	19.69
Exposure group									
M	32.48	11.97	23.08	42.74	15.38	21.37	27.35	11.11	29.06
SD	20.01	9.58	16.01	21.68	14.73	16.64	20.09	11.11	18.45
Round Two									
Instructed group									
M	44.44	20.63	22.22	31.75	17.46	15.87	26.98	12.7	25.4
SD	33.33	24.37	23.13	26	25.55	19.09	22.09	18.62	29.89
Exposure group									
M	34.92	15.87	33.33	34.92	14.29	3.17	34.92	20.63	22.22
SD	19.7	16.8	15.71	26	18.94	15.33	23.51	18.62	28.69

Concerning the general trends seen in Tables 5.1 and 5.2, several points can be made. First, Remember scores for the instruction group are consistently higher for both grammar and vocabulary at testing time 1 (T1) across both data collection rounds, and are generally higher across all testing times. This, as will be discussed below, along with other data, lends support to the main hypotheses of this project. Second, concerning the comparison of grammar noticing and vocabulary noticing, two trends can be seen from this data. Grammar noticing scores are consistently higher for the instruction group than the exposure group, while scores for vocabulary noticing show more variable results between the instruction and exposure group. Also, Remember responses in general are consistently higher for vocabulary noticing than for grammar noticing across both the instruction and exposure groups. This, and other data which will be presented below, suggest important differences between the noticing of grammar and vocabulary. Third, rather consistent differences can be found between the

scores for Remember, Know and Guess responses. Remember responses are generally consistently higher than Know and Guess responses at each testing time for both grammar and vocabulary noticing for both the instructed and exposure group. Also, there is a trend where Know and Guess responses are generally more accurate for the instructed group than for the exposure group. Issues relating to all three response types will be presented and discussed in the chapter on learning and in the general discussion.

These points and others will be presented and discussed in detail in the various sections below in more detail. What immediately follows is the presentation and discussion of the results as they directly relate to the hypotheses presented in Chapter 3. Also, note that in the following sections, along with other data, data presented in Tables 5.1 and 5.2 will be presented again in slightly different tabular format in order to be more comprehensible.

5.2. RESULTS CONCERNING THE EFFECTS OF INSTRUCTION VS. EXPOSURE ON THE NOTICING OF GRAMMAR

Round One and Two results for the test of noticing grammar are shown in Table 5.3. In Round One, scores for the instructed group are consistently higher across all three testing times, though only scores at testing time one (T1) are shown to be significantly different (t = 3.050, p = .003). Across all testing times and both rounds, the highest scores are obtained by the instructed group at T1.

Results for Round Two show that the instructed group achieved higher scores for noticing at T1 and T2, while at T3 the exposure group had higher scores. Statistical testing showed that none of the differences in Round Two were significant for grammar. Here again, as in Round One, the instructed group at T1 has the highest score across both groups and the three testing times. These results are presented graphically in Figures 5.1 and 5.2.

Table 5.3. Noticing scores for target grammar for the instructed and exposure groups.

	T1		T2		T3	
	Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
Round 1						
M	38.46	10.77	23.08	20	24.62	10.77
SD	23.75	22.53	17.97	25.82	26.02	22.53
p	.006*		0.727		0.16	
Round 2						
M	42.86	28.57	20	5.71	17.14	20
SD	26.9	27.95	16.33	19.02	17.99	23.09
p	0.053		0.158		.696	

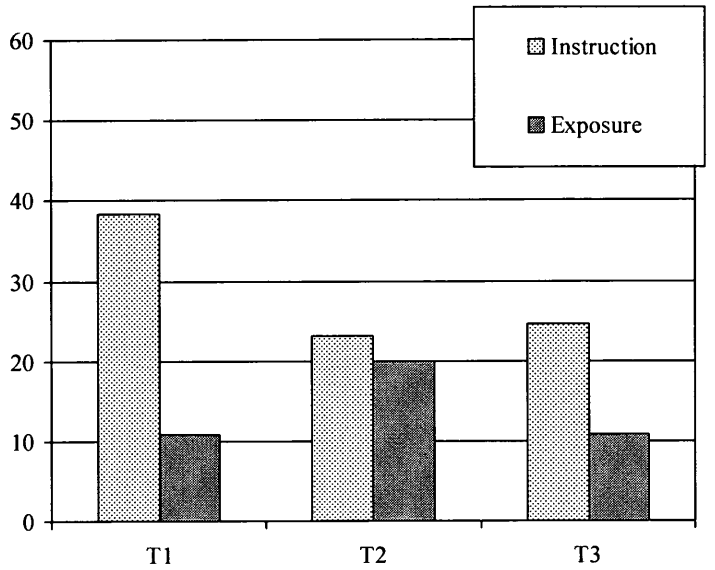


Figure 5.1. Round One noticing scores for target item grammar for the instruction and exposure groups.

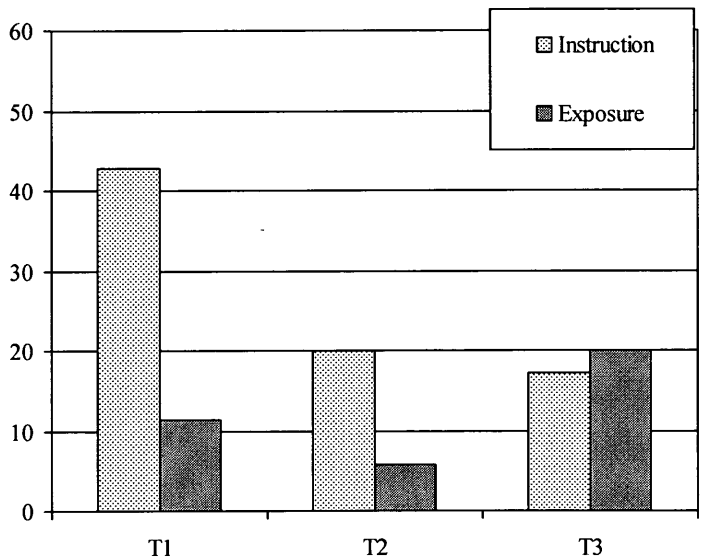


Figure 5.2. Round Two noticing scores for target item grammar noticing for the instruction and exposure groups.

5.3. RESULTS CONCERNING THE EFFECTS OF INSTRUCTION VS. EXPOSURE ON THE NOTICING OF VOCABULARY

Round One results for vocabulary are shown in Table 5.4. The instructed group achieved higher accuracy rates for noticing at T1 and T3, but the exposure group had higher scores at T2. Statistical testing shows that results were statistically significant on at T1 ($t = 2.301$ $p = .015$). As before, the highest score of all three testing times was achieved by the instructed group at T1.

Table 5.4. Noticing scores for target vocabulary for the instruction and exposure groups.

	T1		T2		T3	
	Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
Round 1						
M	54.70	32.48	37.61	42.74	43.59	27.35
SD	28.50	20.01	21.53	21.68	26.24	20.09
p	0.03		0.551		0.089	
Round. 2						
M	44.44	34.92	31.75	34.92	26.98	34.92
SD	33.33	19.7	26	26	22.09	23.51
p	. 527		0.823		0.68	

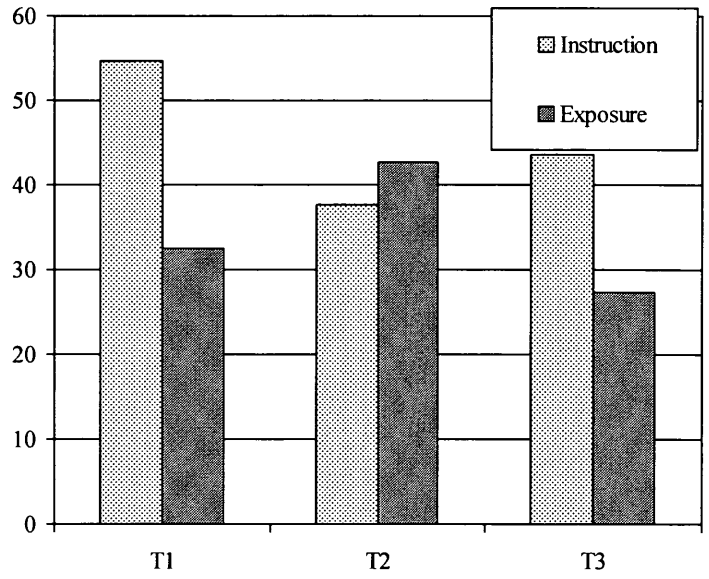


Figure 5.3. Round One noticing scores for target item vocabulary for the instruction and exposure groups.

Round Two results show that the instructed group achieved higher scores at T1, again, as in the previous results, the highest score across all three testing times, but lower scores than the exposure group at T2 and T3. Statistical testing shows that none of the differences between the groups for vocabulary noticing in Round Two were statistically significant.

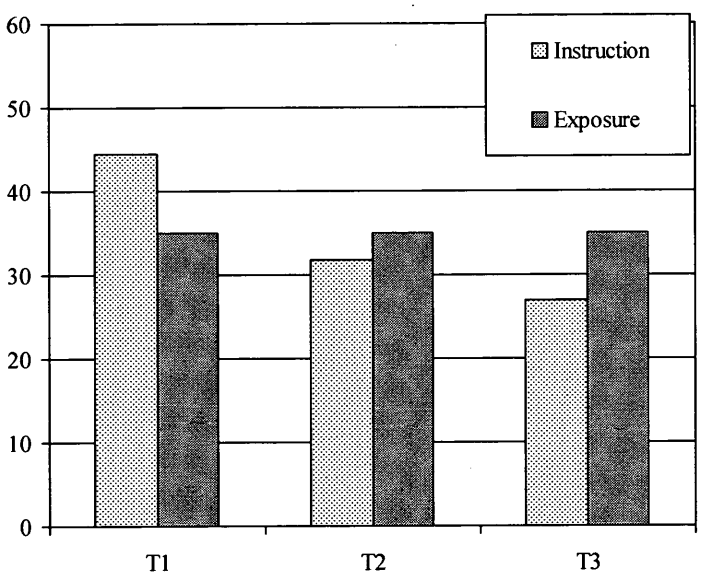


Figure 5.4. Round Two noticing scores for target item vocabulary for the instruction and exposure groups.

5.4. RESULTS CONCERNING THE NOTICING OF INSTRUCTED TARGET ITEM GRAMMAR VS. NON-INSTRUCTED GRAMMAR

Recall that the "non-instructed" items were items which were not the object of instruction or exposure, but simply appeared in the texts used in the test for noticing. The data in this section show within-group comparisons for target and non-instructed group items with the intent of testing whether instruction increases the likelihood of accurate noticing of target items when compared with items which simply appeared in the texts for noticing, that is, non-instructed items.

Table 5.5. Noticing scores for target grammar items, and non-instructed grammar items within the grammar instruction group.

	T1		T2		T3	
	Target	Non-instr.	Target	Non-instr.	Target	Non-instr.
Round 1						
M	38.46	0	23.08	30.77	24.62	25.64
SD	23.75	19.25	17.97	25.32	26.02	19.97
p	.001*		0.453		0.913	
Round 2						
M	42.86	14.29	20	14.29	17.14	0
SD	26.9	26.23	16.33	26.23	17.99	0
p	.009*		0.582		0.045	

These data are presented in Table 5.5. Results for Round One show that while scores are higher for the target items at T1, the scores for non-instructed group items were higher at T2 and T3. Statistical testing shows that only the difference between scores at T1 are significantly different ($t = 4.53$, $p = .001$). Round Two data show a clearer picture. At each testing time, scores on target items are higher than scores on non-instructed items. The difference at T1 is statistically significant ($t = 3.80$, $p = .005$) while the differences at the other two testing times are not significant (at an alpha level of .016). The highest scores for both rounds are found for the target items at T1.

5.5. RESULTS CONCERNING THE NOTICING OF INSTRUCTED TARGET ITEM VOCABULARY VS. NON-INSTRUCTED VOCABULARY

Results comparing target items with non-instructed items are found in Table 5.6. For Round One, it can be seen that at each testing time the scores for the target items are higher than the scores for the uninstructed items, though none of the differences are statistically significant.

For Round Two, results show that, like scores shown in Round One, scores for target vocabulary items are consistently higher than non-instructed items, though none of these differences are statistically significant either. Again, the target items at T1 show the highest scores in both rounds.

Table 5.6. Noticing scores for target vocabulary items, and non-instructed vocabulary items within the vocabulary instruction group.

	T1		T2		T3	
	Target	Non-instr.	Target	Non-instr.	Target	Non-instr.
Round 1						
M	54.7	53.85	37.61	25	43.59	36.54
SD	28.5	26.7	21.53	17.68	26.24	36.25
p	0.717		0.638		0.045	
Round 2						
M	44.44	21.43	31.75	25	26.98	21.43
SD	33.33	39.34	26	28.87	22.09	22.49
p	0.641		0.629		0.592	

5.6. RESULTS OF THE COMPARISON OF TARGET GRAMMAR ITEMS WHICH WERE THE OBJECT OF EXPOSURE WITH TARGET VOCABULARY ITEMS WHICH WERE THE OBJECT OF EXPOSURE

The results presented in Table 5.7 show that for Round One across all three testing times scores on vocabulary were consistently higher than those for grammar. The difference at T1 is statistically significant ($t = 2.597$, $p = .008$), as is the difference at T2 ($t = 2.377$, $p = .012$) while the differences at the other testing times are not significant (again, at an alpha level of .016). Round Two results show the same pattern where vocabulary scores are greater than grammar scores. None of these differences in Round Two are statistically significant. (Note: an analysis of the corresponding instructed items can be found in Section 5.7.3)

Table 5.7. Noticing scores comparing target grammar items which were the object of exposure vs. target vocabulary which were the object of exposure.

	T1		T2		T3	
	Gram.	Voc.	Gram.	Voc.	Gram.	Voc.
Round 1						
M	10.77	32.48	20	42.74	10.77	27.35
SD	22.53	20.01	25.82	21.68	22.53	20.09
p	0.016*		0.023		0.059	
Round 2						
M	11.43	34.92	5.71	34.92	20	34.92
SD	27.95	19.7	19.02	26	23.09	23.51
p	0.094		0.035		0.254	

5.7. DISCUSSION

5.7.1. HYPOTHESES 1 AND 2

Hypotheses 1 and 2 predict the superiority of the instruction group over the exposure group on both the tests of noticing target grammar and target vocabulary. These are the core hypotheses for this study. Two types of data will be discussed in this section, primary data referring to the direct comparison of the two experimental groups on measures of noticing, and secondary data which can lend additional support to the hypotheses.

5.7.1.1. Primary data concerning Hypotheses 1 and 2

Concerning the direct comparison of the experimental groups, though the results vary across the three testing times and the two rounds of data collection, Tables 5.3 and 5.4 and Figures 5.1 through 5.4 show that Hypotheses 1 and 2 can both be seen to be generally supported by the data, with stronger results seen for grammar than vocabulary. Three arguments concerning these data lend support to this claim.

First, statistically significant results, when found, are found in favor of the instructed group and not in favor of the exposure group. This occurred in Round One and Round Two at T1 for both grammar and for vocabulary. These results, though, are the only significant results across the three testing times and both rounds of data collection. This lack of significant results can probably be attributed to the small sample size of 13 participants in each group for Round One and seven in each group in Round Two.

Second, putting aside the issue of statistical significance for the moment, if the general trend in results is looked at, higher scores are generally found for the instructed group over the exposure group. This pattern is particularly strong in the case of the noticing of grammar, where in Round One the instruction group scored better than the exposure group across all testing times, and where in Round Two higher scores are achieved by the instructed group for T1 and T2. For vocabulary, the picture is more mixed. In Round One, higher scores are achieved by the instructed group at T1 and T3, while in Round Two higher scores are found only at T1. One potential reason for the lack of positive results for the instructed group in Round Two has to do with the unusually consistently high scores for vocabulary noticing shown by the exposure group when compared with the results for Round One for the same group, which could have to do with different characteristics of the groups. Tables 4.1 and 4.2, presented previously, show that the grammar instruction/vocabulary exposure group has higher, though not significantly higher scores on all group characteristics except vocabulary pre-test knowledge. A second, and more convincing explanation, one which will be taken up and discussed later in detail, is that it may simply be easier to notice vocabulary than grammar, thus giving the exposure group a greater chance to perform well.

A third piece of evidence in support of Hypotheses 1 and 2 are the data from T1 in both Round One and Two, which show a clear recency effect for noticing. These results show that the instructed group consistently outperformed the exposure group at the first testing times, with two of the differences being statistically significant. Furthermore, in all cases — for both grammar and vocabulary for both data collection rounds — these scores at T1 were the

highest scores achieved across all testing times. A strong initial effect and later decline in results such as this is exactly what would be expected when measuring the retention of memories, and this decline in scores for Remembering is what has been found in other studies using the Remember/Know paradigm (Gardiner & Java, 1991). What this suggests, then, is that after an initial period of time during which instruction may lead to noticing, results fall off to a level where participants who were not instructed have an equal chance of noticing items — particularly vocabulary — when compared to the instructed group.

It is instructive here to compare the scores for the instruction group to those of the exposure group, where a recency effect cannot be found. For grammar in Round One, the highest score for exposure occurs at T2, while in Round Two it appears at T3. For Round One vocabulary the highest score for exposure is at T2, and for Round Two, a slightly different effect is found where the same score was achieved by the exposure group at all three testing times. This comparison between the instruction and exposure groups then suggests that there is a qualitative difference between the two groups, one, the instruction group, following the pattern that might be expected when testing the retention of more elaborately processed materials, and the other, the exposure group, showing no recency effect and more variable results as could be expected for memories created through shallower and less elaborate processing.

The evidence described immediately above is also useful in discussing whether or not the results achieved can be explained solely on the basis of an item effect. Recall that at each testing time a slightly different group of items was selected for testing so as to reduce the possibility of a testing effect. This opens up the possibility that at each testing time, and, of course, across the two data collection rounds, results could simply be explained due to the testing of different items. The evidence of a consistent pattern of results where the instructed group scores better and achieves the highest scores of each round at T1 can discount this possibility to some degree, though it is still possible that an item effect could have had a role in the varied results found at later testing times. A detailed look at performance on individual items will be taken up in Chapter 8.

Thus, considering the three points above, there is rather strong support for Hypothesis 1, concerning the beneficial effects of instruction on the noticing of grammar. Hypothesis 2, concerning the positive influence of instruction on the noticing of vocabulary, is also supported, though not as strongly as Hypothesis 1. Furthermore, two other issues were raised, the probability that vocabulary is easier to notice than grammar, and the possibility of an item effect influencing results. The issue of vocabulary being easier to notice will be taken up below in the following sections where further evidence in support of Hypotheses 1 and 2 will be presented.

Another issue should be raised here concerning an unusual occurrence in the Round One data for both the grammar and vocabulary. In Figure 5.5 and Figure 5.6, Round One grammar and vocabulary noticing scores are presented in a line graph format, which allows the following pattern to be seen more easily: for both grammar and vocabulary, instructed scores drop off at T2 and then recover slightly at T3, while exposure scores for grammar and vocabulary undergo the opposite effect of spiking slightly at T2 and returning to a lower level at T3.

Although the results are not statistically significant across all the testing times, the issue is still raised of how to explain these unusual results.

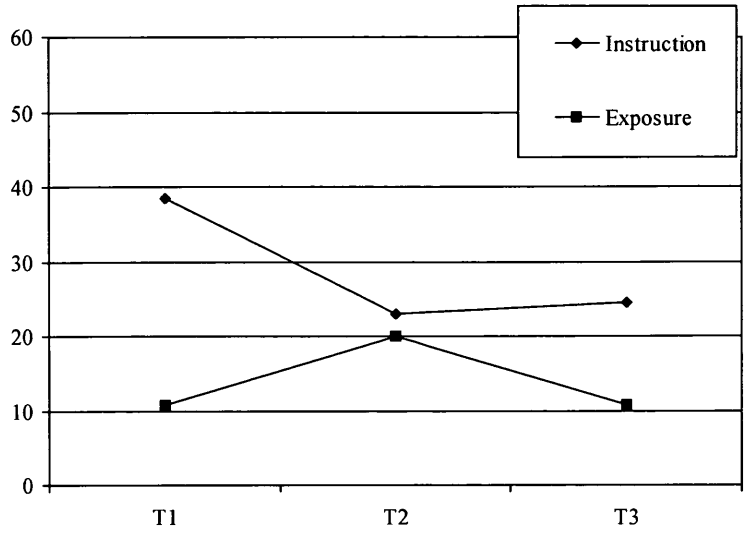


Figure 5.5. Round One target item grammar noticing for the instruction and exposure groups.

One likely, although speculative, possibility is that these results are due to participants' test taking strategies. Recall that each group of participants served the dual purpose of being the instructed group in one area, say grammar, and also serving as the exposure group in the opposite area, vocabulary in this example. That is, the instructed group for grammar is the same group which produced the scores for the exposure group in vocabulary, and vice versa. Looking at the results in Figures 5.5 and 5.6, this means that the unusual pattern cannot simply be explained by situational factors in the classroom — say, a lack of attention, or a disturbance — which would undermine scores in general by dividing attention. If this were the case, then a general drop off in scores for both instruction and exposure for the same group would be expected, but what is found is an increase in scores for exposure and a decrease for scores for instruction. Although general situational factors seem an unlikely explanation, learners' test taking strategies can be used to account for these data in the following way. Each group was instructed in only one area, but when they experienced the first test, they were tested on both the noticing and learning of grammar and vocabulary. Anecdotally, it could be observed at the time of testing that participants were surprised at being tested on things in which they had not been instructed. At the second testing time, then it is possible that participants used a strategy of directing their attention during the reading of the passage used in the noticing test to the general area in which they had no instruction in an attempt to bolster their test scores. This would fit in well with the general profile of the participants as high-achieving test-savvy students existing in a competitive educational environment. If this

account is correct, then, it appears that participants' strategies at the time of testing can have an effect on the level of noticing. This issue will be taken up again in later sections.

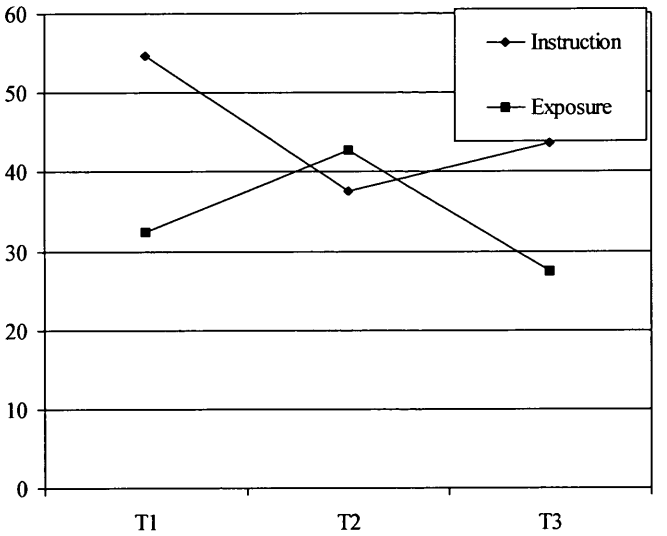


Figure 5.6. Round One target item vocabulary noticing for the instruction and exposure groups.

Finally, it should be noted that the strong results achieved for instruction concerning grammar over vocabulary — five out of six higher scores across all testing times and both data collection rounds support instruction for grammar compared with three out of six supporting instruction for vocabulary — point to the possibility that instruction is more helpful for noticing grammar than for noticing vocabulary. This is a point which will be taken up later as well.

5.7.1.2. Secondary data concerning Hypotheses 1 and 2

One source of data which can lend strength to the support of Hypotheses 1 and 2 are the scores for the so-called "+target/–text" items. These are the items which were the object of instruction or exposure, but which were purposefully left out of the text used for the test of noticing. On the list used in the noticing test, these items were paired with items which had actually appeared in the text used on the noticing test. Thus, these items were used as distractors to test whether participants in the instructed group were marking items as noticed simply because they had been instructed in them, or because the items had appeared in the text. Thus, the ability to accurately reject the +target/–text items (that is, the incorrect answer) would be evidence that instructed learners are accurately noticing target items in the texts.

Table 5.8. Noticing scores indicating the accuracy of rejecting instructed grammar items which were not in the text of the noticing (+target/–text items) for the instruction and exposure groups.

	T1		T2		T3	
	Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
Round 1						
M	15.38	0	0	0	30.77	38.46
SD	55.47	40.82	40.82	0	48.04	50.64
p	0.34		1		0.75	
Round 2						
M	0	14.29	–4.29	14.29	14.29	0
SD	0	37.8	37.8	37.8	37.8	0
p	0.17		0.09		0.17	

Table 5.9. Noticing scores indicating the accuracy of rejecting instructed vocabulary items which were not in the text of the noticing test (+target/–text items) for the instruction and exposure groups.

	T1		T2		T3	
	Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
Round 1						
M	12.82	0	33.33	28.21	17.95	23.08
SD	25.6	13.61	33.33	32.9	17.3	28.5
p	0.21		0.72		0.55	
Round 2						
M	28.57	19.05	4.76	14.29	0	0
SD	29.99	26.23	12.6	17.82	0	0
p	0.27		0.14			

Tables 5.8 and 5.9 present these data for grammar and vocabulary for Rounds One and Two. It is important to note that these scores represent the accuracy with which participants have *rejected* the +target/-text items and chosen the item which had actually appeared in the text. Thus, the higher the score, the more accurately the group was able to supply a Remember response in rejecting the +target/–text item. It is also important to note that there are very small numbers of items being tested here, just a single grammar item and three vocabulary items made up this group, thus it is not possible to draw strong conclusions from these data. Finally, it is important to mention scores of zero could show that no Remember response was given for that particular group of items, or that an equal number of correct and incorrect answers were given, as can be seen reflected in the standard deviations. Although none of the differences between the groups are statistically significant, one rather clear conclusion can be drawn from the data: the instructed group is rarely lured into giving a Remember response for an item simply because this item was the object of instruction and ap-

peared on the list of items for the noticing test. The instructed group does not display the very large negative scores that would be expected if answers were being given on the noticing test largely on the basis of familiarity of target items through instruction alone. These data thus support the validity of the scores achieved by the instructed group and reported in the previous sections.

A second source of data which can also support the conclusions made in the previous sections are the data from the retrospections which were collected at T1 for both Round One and Round Two. Tables 5.10 and 5.11 present a categorization of participants' retrospections concerning their reasons for giving Remember responses. Participants' responses were placed into six categories. The basic categories were established during the pilot project for this research and then were expanded when these current data were analyzed. The first three categories, "studied before", "unknown or unusual", or "position on page", are self-explanatory. Category 4, "formed key part of the story", was included due to the fact that participants would oftentimes remember a word, or, much less frequently, a grammar item, due to its position or function in the narrative itself and could, in their retrospections, retell that part of the story using the particular word or grammatical construction. Category 5 is similar to 4, although these responses appeared to be more related to the items themselves. That is, the participants could repeat back the exact line or phrase from the story which contained the target item, but this was not contextualized in the narrative. The final category, number 6, was maintained for those participants who claimed they had experienced the item in the text, but were unable to describe exactly what this experience was. This was most often expressed in terms of simply being able to remember "seeing the word in the text". This lack of specificity of the context of remembering does make the validity of these Remember responses suspect, but they were qualitatively different than the Know responses in being able to provide some reference to the context of "seeing it", whereas Know responses are often accompanied by an inability to explain anything contextual about the reasons for the feelings of familiarity of the item (Gardiner, 1998:7).

Table 5.10. Frequency of reasons given during retrospections for Remember responses for Round One target items at T1.

Reason for Remember response:	Grammar		Vocabulary	
	Inst.	Exp.	Inst.	Exp.
1. Studied before	7	0	30	5
2. Unknown or unusual	13	11	1	13
3. Position on page	0	0	2	1
4. Formed key part of story	0	0	5	5
5. Can repeat back line of text with item in it.	0	0	5	7
6. Remember seeing/visualizing it in the text.	7	1	16	10

The data in Tables 5.10 and 5.11 provide evidence that at least one of the most frequent reasons that the instruction group chose Remember responses was because of the effects of instruction: when encountering the items in the text, many participants reported recalling at

that time that they had been instructed in that particular item. The "studied before" category has the largest number of responses everywhere except for Round One grammar, where the "unknown or unusual" category received the most frequent responses. It was unclear without further probing of participants' answers if, for the instructed group, underlying the responses other than "studied before" was also a conscious experience relating to remembering the items as having been instructed. Nevertheless, a fairly clear effect can be shown where items in input become salient for the instructed group due to previous instruction, and this gives support to the claim of instruction having an impact on noticing at input.

Table 5.11. Frequency of reasons given during retrospections for Remember responses for Round Two target items at T1.

Reason for Remember response:	Grammar		Vocabulary	
	Inst.	Exp.	Inst.	Exp.
1. Studied before	10	0	12	2
2. Unknown or unusual	2	4	1	15
3. Position on page	0	0	0	1
4. Formed key part of story	0	0	4	0
5. Can repeat back line of text with item in it.	0	0	4	1
6. Remember seeing/visualizing it in the text.	1	3	6	3

5.7.2. HYPOTHESES 3 AND 4

These hypotheses test within-group differences concerning comparison of the noticing of target group items with items which naturally occurred in the texts used in the test for noticing, the non-instructed items. As such, along with the main hypotheses, 1 and 2, Hypotheses 3 and 4 can possibly add supporting evidence for a general picture supporting the efficacy of instruction in causing noticing.

Hypothesis 3, that higher noticing scores will be achieved for target group grammar items than the non-instructed group grammar items, is generally supported by the data, as was previously shown in Table 5.5. The significant differences which occur between target and non-instructed grammar items are in favor of the hypothesis, both in Round One and Round Two at T1, show that instructed items have greater saliency for participants than other grammar items appearing in the text used for noticing. This effect at T1, as mentioned above, is most likely due to a recency effect, and it is important that this effect still holds when compared to non-instructed items. Finally, again ignoring the issue of statistical significance for the moment, it can be seen that across the two rounds of data collection, higher scores are achieved for instructed items in four out of six cases. Thus, looking across all the data, the results can be seen to be supportive of the hypothesis, though the data are not conclusive.

Hypothesis 4 predicts that target group vocabulary items will be noticed to a greater degree than the non-instructed group vocabulary, and the data are generally supportive of the hypothesis, as was seen in Table 5.6. Although there are no significant differences between scores at any of the testing times across both data collection rounds, in each case the general

trend is in favor of the target items over non-instructed items. Again, as in the discussion of the previous hypotheses, the highest scores which are achieved are for instructed items at T1 in each round of data collection. Thus, the data are supportive of the hypothesis, though without conclusive statistically significant results.

Looking across both sets of data for grammar and vocabulary, it is possible to suggest that instruction gives an advantage to noticing items in input, most particularly immediately after instruction, when compared to non-instructed items. Thus, these data support the general picture for the beneficial effects of instruction on noticing, but Hypotheses 3 and 4 are also important for another reason. Given the general support for the hypotheses, it is not possible to discount the stronger effects found for Hypotheses 1 and 2 by claiming that the instructed group was simply noticing all items which appeared in the texts for noticing under, perhaps, a generalized influence of instruction where sensitivity to grammar or vocabulary was heightened. If this were the case, there should not be a difference between scores of instructed items and non-instructed items.

It is still possible, though, that the items chosen for instruction were simply more salient items than those non-instructed items chosen from the texts used in the tests for noticing. Evidence on this point is mixed, as can be seen by looking at the data for target items and the non-instructed items produced by the exposure group. Scores for target item and non-instructed item grammar and vocabulary items produced by the exposure group can be found in Tables 5.12 and 5.13. Though none of the comparisons are statistically significant, a clear pattern emerges. In Round One, for both grammar and vocabulary, scores on non-instructed items are higher than scores on target group items. This would suggest that in this round, the instructed grammar and vocabulary items used were no more salient than the items chosen for the unexposed group. On the other hand, for the exposure group in Round Two, target group items consistently receive higher scores for both grammar and vocabulary. This clear division between the two data collection rounds suggests that there may have been a difference between the target items used.

Table 5.12. Noticing scores for target grammar items and non-instructed grammar items within the exposure group.

	T1		T2		T3	
	Target	Non-instr.	Target	Non-instr.	Target	Non-instr.
Round 1						
M	10.77	15.38	20	30.77	10.77	12.82
SD	22.53	29.24	25.82	16.45	22.53	16.88
p	0.618		0.106		0.827	
Round 2						
M	11.43	4.76	5.71	4.76	20	0
SD	27.95	12.6	19.02	12.6	23.09	19.25
p	0.641		0.899		0.166	

Table 5.13. Noticing scores for target vocabulary items and non-instructed vocabulary items within the exposure group.

	T1		T2		T3	
	Target	Non-instr.	Target	Non-instr.	Target	Non-instr.
Round 1						
M	32.48	36.54	42.74	46.15	27.35	44.23
SD	20.01	41.6	21.68	28.59	20.09	30.88
p	0.717		0.683		0.045	
Round 2						
M	34.92	28.57	34.92	25	34.92	32.14
SD	19.7	26.73	26	32.27	23.51	40.09
p	0.583		0.387		0.803	

5.7.3. HYPOTHESIS 5

Hypothesis 5 predicted greater scores for the noticing of vocabulary than grammar for the exposure group. This hypothesis is strongly supported by the data found in Table 5.7. Across both rounds of data collection, scores are markedly higher for vocabulary than grammar, though significant differences only occur in Round One at T1 and T2

Table 5.14. Noticing scores comparing target item grammar which was the object of instruction vs. target item vocabulary which was the object of instruction.

	T1		T2		T3	
	Gram.	Voc.	Gram.	Voc.	Gram.	Voc.
Round 1						
M	38.46	54.7	23.08	37.61	24.62	43.59
SD	23.75	28.5	17.97	21.53	26.02	26.24
p	0.128		0.074		0.076	
Round 2						
M	42.86	44.44	20	31.75	17.14	26.98
SD	26.9	33.33	16.33	26	17.99	22.09
p	0.924		0.331		0.397	

The rationale for restricting the hypothesis to target items which were the object of exposure was that it was not possible to predict what differential effect instruction would have on grammar and vocabulary. That is, although there was reason to believe that vocabulary is inherently more likely to be noticed due to its greater surface level saliency while grammar is more likely to be more difficult to notice due to its more abstract and complex nature, it was not possible to predict whether or not instruction could level out these differences or whether the expected differences would occur even with instruction. Post-hoc, though, it is possible to make this comparison, and, as can be seen from Table 5.14, even with instruction,

scores for vocabulary are consistently higher than for grammar. Statistical testing showed no significant differences with these data, yet still a general trend can be seen which is suggestive of an advantage for vocabulary noticing in general.

Thus, Hypothesis 5 receives strong support from a variety of sources which show that, under these experimental conditions, vocabulary can be noticed to a greater degree than grammar. Support for this hypothesis is also important for interpreting the generally superior results for grammar noticing in the instructed condition; Hypothesis 5 shows that vocabulary is naturally more noticeable, thus lessening the potential effects of instruction between the exposure and instruction group.

5.7.4. RESULTS USING REMEMBER + KNOW RESPONSES TO DEFINE NOTICING

This study is based on the theoretically-backed assumption that Remember responses represent noticing. Yet, it is appropriate to again point out here that there are different definitions which would support the claim that noticing can be related to both the states of remembering and knowing. The key distinction lies between the definitions of noticing proposed by Schmidt (1990), on which this study is based, and Tomlin & Villa (1994). Schmidt's definition of noticing, expressed in Tomlin & Villa's terminology, requires detection of stimuli plus conscious awareness of them, while Tomlin & Villa's minimal definition simply involves detection without necessarily involving awareness, yet allowing for the possibility of awareness in the definition. Thus Tomlin & Villa have a wider definition of noticing. It might be possible, then, to equate noticing, as Tomlin & Villa would see it, to both the Remember and Know responses, as Know responses rely only on a strong feeling that an item has been encountered, without a memory of having been conscious of the item at the time of encoding. That is, this could be an example of detection without awareness. Thus, it is appropriate to look at combined Remember and Know scores for the main hypotheses, 1 and 2, to see if the results are different than those calculated just using remember scores.

Table 5.15 shows the results for grammar and vocabulary across both data collection rounds. Concerning grammar, a comparison of Table 5.15 with Table 5.3, which presented the scores for Remember responses on grammar, shows that Remember + Know data present slightly stronger results than just using the Remember responses. In the Remember + Know data all differences are in favor of the instruction group, rather than five out of the six across both rounds for the Remember responses, and the differences at T1 for both Round One and Round Two are statistically significant, rather than the differences just at Round One T1.

The results for vocabulary are similar. In a comparison with Table 5.4, it can be seen that whereas the instruction group scores higher in three out of the six cases across both data collection rounds concerning the Remember data alone when combining Remember and Know data together, the instruction groups scores higher than the exposure group in four cases and equal to the exposure group in one case. Also, where no significant differences are found between groups concerning the Remember data alone, the differences for Round One T1 and T2 are found to be significant using the Remember + Know data.

Table 5.15. Noticing scores for target grammar and vocabulary for the instructed and exposure groups combining both Remember and Know scores.

	T1		T2		T3	
	Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
Grammar						
Round 1						
M	38.46	4.62	24.62	21.54	30.77	10.77
SD	29.41	29.61	18.54	38.7	36.16	23.97
p	.005*		0.79		0.11	
Round 2						
M	68.57	8.57	25.71	22.86	25.71	22.86
SD	27.95	30.24	25.07	31.47	27.6	23.52
p	.001*		0.427		0.432	
Vocabulary						
Round 1						
M	74.36	44.44	51.28	58.12	63.25	38.46
SD	15.96	18.14	21.53	17.66	23.3	19.04
p	.001*		0.39		.007*	
Round 2						
M	65.08	50.79	49.21	49.21	55.56	39.68
SD	18.62	21.14	14.14	27.73	20.29	27.86
p	0.1		0.5		0.123	

Thus combining Remember and Know responses makes the results stronger to some degree. Yet, while this is an interesting option for expanding the definition of noticing, the benefit of limiting the definition of noticing to Remember responses is that it requires the conscious processing which occurs in working memory, which is hypothesized to be a key factor in converting input to intake (Robinson, 1995b). Nevertheless, these results are interesting, and further differences between Remember and Remember + Know responses will also be considered in the following chapter.

5.8. SUMMARY

Data concerning the noticing of grammar and vocabulary were presented in this chapter which generally showed positive effects of the role of instruction on noticing in input. It was noted that while, overall, statistically non-significant results were found, generally strong patterns can be found in the data which give support to the hypotheses. For Hypotheses 1 and 2, predicting superior results for the instruction group over the exposure group on the noticing of grammar and vocabulary, respectively, results showed largely positive results with stronger results for the noticing of grammar. Hypotheses 3 and 4 concerned the comparison of target

vocabulary and grammar items, respectively, which were either the object of instruction or exposure with non-instructed items which occurred naturally in the texts for noticing. Again, these predictions were generally supported by the data. Hypothesis 5 predicted greater noticing of target vocabulary than target grammar in the exposure condition, and strong results were found for this hypothesis in the data. Finally, the issue of expanding the definition used in noticing to include Remember + Know data was considered, and it was shown that by doing this the general pattern established by the Remember responses could be strengthened.

CHAPTER 6. RESULTS AND DISCUSSION CONCERNING NOTICING AND LEARNING

The previous section discussed noticing in general and the relationship between noticing and instruction. It was demonstrated that an effect of instruction could be found where instruction positively affects noticing when compared to exposure. This section will first look at the effect of instruction on learning, and then in the following section the issue of noticing will be returned to in order to explore the possible relationship, or co-occurrence, between noticing and learning.

It is important to recall at the beginning of this section that learning scores were calculated in two different ways. For analysis of individual data involving the plotting of individual scores of noticing against individual scores on learning, adjustments were made in participants' scores to account for prior knowledge of vocabulary and grammar items, as was described in section 4.6.4 in the methodology chapter. This was done so as not to bias results in favor of learning. For comparisons between groups, as will be done immediately below, no adjustments for prior knowledge were made, as groups were shown to be equivalent in terms of prior knowledge in pre-testing. The removal of items from an already small group of total items scored by a small group of participants can reduce the reliability of the data, and for this reason it was decided to report the non-adjusted scores for between-group comparisons.

6.1. RESULTS FOR LEARNING

Results will be presented first for grammar and then for vocabulary in order to establish that learning did indeed occur, and to demonstrate the differential rates of learning that occurred due to the effects of instruction or exposure. The testing of grammar learning consisted of two tasks, a sentence completion task and a grammaticality judgement task. Scores for the sentence completion task for Round One and Two are reported in Table 6.1. As can be seen from the table, instructed learners consistently scored higher than the exposure group. Statistical testing shows that the differences at T1 and T2 for Round One were significantly different ($t = 5.085, p = .001$; $t = 3.403, p = .001$, respectively), while the difference at T3 was not. In Round Two, the differences at each time were significant ($t = 5.818, p = .001$; $t = 6.832, p = .001$; $t = -3.97, p = .001$).

Table 6.1. Learning scores showing percentage correct for grammar on the sentence completion task for Round One and Two for both instruction and exposure groups.

	T1		T2		T3	
	Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
Round 1						
M	68.33	23.33	51.54	16.79	54.23	26.28
SD	23.29	13.64	25.53	14.63	33.96	20.7
p	.001*		.002*		0.122	
Round 2						
M	66.43	14.29	81.43	17.14	82.14	31.43
SD	16.51	19.02	14.35	21.38	18.22	27.95
p	.001*		.001*		.002*	

A similar picture of the data for grammaticality judgements is found in Table 6.2. Here, the instructed learners again outperform the exposure group at each point. The difference at Round One T1 was statistically significant ($t = 3.937$, $p = .001$), as were all the results in Round Two ($t = 6.339$, $p = .001$; $t = 3.308$, $p = .006$; $t = 7.273$, $p = .001$). It is important to note that the very high scores for the grammaticality judgement task are due, in part, to the fact that a score of 50% could reflect chance performance.

Table 6.2. Learning scores showing percentage correct for grammar on the grammaticality judgement task for Round One and Two for both instruction and exposure groups.

	T1		T2		T3	
	Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
Round 1						
M	72.5	53.97	64.62	59.42	65.77	58.65
SD	12.46	11.63	12.78	15.25	10.92	15.33
p	.001*		0.312		0.07	
Round 2						
M	74.29	37.14	83.57	50	78.21	45.71
SD	10.28	7.56	10.29	25.17	8.26	9.76
p	.001*		.006*		.001*	

The scores for vocabulary learning for both Rounds One and Two are presented in Table 6.3. Again, the data for both rounds show that the instructed group consistently achieved higher scores than the exposure group across both rounds with statistically significant results found in Round One at T1 ($t = 4.536$, $p = .001$), and significant differences found at each testing time in Round Two ($t = 6.339$, $p = .001$; $t = 3.308$, $p = .003$; $t = 7.273$, $p = .001$).

The overall, view for learning, then, is one where instruction consistently allows participants to achieve higher scores on both grammar and vocabulary across all three testing times.

Finally, while comparisons across time are not possible due to the different composition of the group of target items used at each time, it is possible to speculate from the data here that learning results are surprisingly stable and durable. Even at T3, a testing time just over six weeks from the time of instruction for Round One and three weeks for Round Two, participants score very similar to, and sometimes better than, scores from T1 and T2.

Table 6.3. Learning scores showing percentage correct for vocabulary for Round One and Two for both instructed and exposure groups.

	T1		T2		T3	
	Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
Round 1						
M	27.27	7.91	20.95	12.19	25.8	6.05
SD	19.41	10.64	20.34	16.55	21.01	8.95
p	0.001*		0.53		0.055	
Round 2						
M	35.05	11.44	35.78	6.61	31.94	4.23
SD	17.71	11.6	17.58	7.08	17.53	4.98
p	.011*		.002*		*.0001	

6.2. RESULTS PLOTTING NOTICING AGAINST LEARNING

The correlations found in Table 6.4 can serve to introduce the data concerning the relationship, or co-occurrence of noticing and learning. Here, correlations are found between noticing scores and the two tests of grammar learning (the sentence completion task and the grammaticality judgement task) and also the test of vocabulary learning. Following this, data specific to the issue of whether learning can occur without noticing will be presented. Note that the data presented here for learning and noticing are slightly different than the data presented above, as scores have been adjusted to account for prior learning. It is also important to note that both the instructed and uninstructed groups are included together for the purpose of this analysis. This is done in order to look at an overall relationship between noticing and learning. Furthermore, both data for Remember responses and Know responses are presented here, though the correlations with the Remember responses will be discussed first and most exhaustively.

It is necessary to point out here as well that these correlations are generally being computed for very small groups and therefore may be unreliable. Furthermore, it is quite possible that there are other factors, such as aptitude, which may be affecting these correlations. Thus, these factors must be taken into account when looking at these data.

Concerning grammar noticing and learning, the correlations between sentence completion and noticing, using Remember scores, for Round One show low moderate to low positive

correlations, gradually declining from T1 to T3. The data for Round Two show similar results, although the correlations for T1 and T2 are stronger than those found in Round One, and there is a sharper drop off to a low negative correlation at T3. Statistically significant results (at the .016 alpha level, again due to multiple testing of the same variable across different testing times) are found only in Round Two at T1.

Table 6.4. Correlations for learning by Remember and Know noticing responses for target items in Round One and Two.

	T1		T2		T3	
	R	K	R	K	R	K
Round One						
Grammar noticing x sentence completion task	0.35	0.19	0.26	0.07	0.19	–.18
p	0.036	0.181	0.101	0.36	0.175	0.195
Grammar noticing x grammaticality judgement task	0.22	0.3	-0.03	0.3	0.31	–.37
p	0.138	0.066	0.449	0.26	0.307	0.03
Vocabulary noticing x vocabulary learning	0.6	–.15	0.41	–.13	0.57	–.06
p	.001*	0.231	0.018	0.226	.001*	0.382
Round Two						
Grammar noticing x sentence completion task	0.61	0.73	0.49	0.01	–.13	–.23
p	.001*	.001*	0.037	0.493	0.316	0.21
Grammar noticing x grammaticality judgement task	0.56	0.81	0.58	0.03	0.14	0.3
p	0.018	.001*	.014*	0.458	0.316	0.302
Vocabulary noticing x vocabulary learning	0.56	–.14	0.28	–.13	0.13	–.15
p	0.019	0.312	0.146	0.334	0.325	0.304

Concerning the data for correlations between noticing and the grammaticality judgement task, again, for the Remember responses, the data in Round One show a slightly different result than those for the sentence completion task. The correlations are generally lower, and the pattern, rather than showing a gradual decline across testing times, shows a steep drop off from T1 to T2 and then a recovery at T3 to a score higher than at T1. In Round Two, correlations are very similar to the ones found for noticing and the sentence completion task in that same Round with statistically significant results are at T2.

In the vocabulary results, it can be seen that in Round One there is a moderate positive correlation between noticing and learning found at each testing time, with the highest correlations found at T1 and T3. These two correlations are the only statistically significant cor-

relations in this round. A similar, though weaker, pattern can be found in the Round Two vocabulary and noticing data, though in Round Two none of the correlations are statistically significant for vocabulary.

Overall, from the data presented in the correlations, a general picture emerges of good moderate correlations at T1, although in Round One weaker scores are achieved there, with a gradual decline to T3, excepting the grammaticality judgement by noticing correlation in Round One, T3, where there is a large increase in the size of the correlation. Again, though, it must be pointed out that these data have low reliability, and it is not possible to draw strong conclusions from them.

Another, though related, way of looking at these data, which is more directly pertinent to Hypotheses 6 and 7 (which predict that there will be no high learning scores without corresponding high noticing scores), is to graphically plot the data presented immediately above to see the distribution of scores across the possibilities of high and low noticing and high and low scores for achievement on tests of learning.

A key issue here is making a determination as to what can be considered "high" and "low" for noticing and learning, and there are few general guidelines available for making this determination. In one study looking at a relationship between the critical period and aptitude (DeKeyser, 2000), a cutoff point for high aptitude was established by looking for a point which would differentiate a group of scores higher than the mean, and yet still large enough for statistical analysis. This method of consulting the distribution of the data in making a separate determination for the cutoff point for each of the correlations noted above would be preferred, although in the case of this present study, the large amount of data being analyzed would mean making a total of 18 separate decisions of this type considering all three testing times across both data collection rounds. Given the arbitrary nature of decisions of this type, it is preferable to make a single decision used consistently across all the data rather than 18 separate decisions. Thus, a cutoff point for high noticing and high learning was set at one-half standard deviation above the mean. This point appears high enough to allow for the differentiation of a restricted group of scores higher than the mean, and yet low enough to allow a good possibility of finding scores in this range.

A sampling of figures showing noticing scores and grammar and vocabulary scores at T1 for Round One and Two are presented below. The four solid lines on each figure represent the cutoff point of .5 *SD* above or below the mean for noticing and learning. Each data point represents an individual's noticing score and learning score. The issue of low noticing and learning will be addressed in later sections.

Figures 6.1 and 6.2 present data for vocabulary noticing and learning for Rounds One and Two at T1. As can be seen, a relatively small number of participants achieved both high scores for noticing and learning. At the same time, a small number of participants managed to achieve high learning without demonstrating high noticing.

This same pattern can be seen in the scores for grammar noticing as plotted against the results from the sentence completion task for Round One and Round Two at T1, as shown in Figures 6.3 and 6.4.

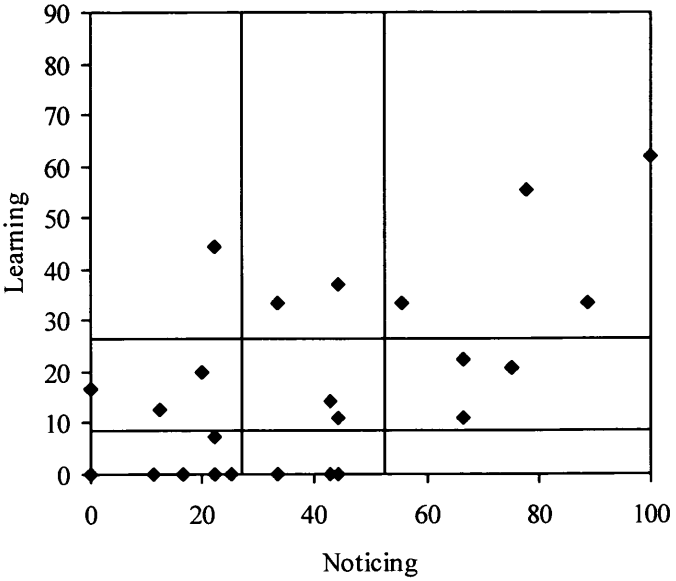


Figure 6.1. Scatter diagram for Round One vocabulary noticing of target items and vocabulary learning at T1, (r=.60).[2]

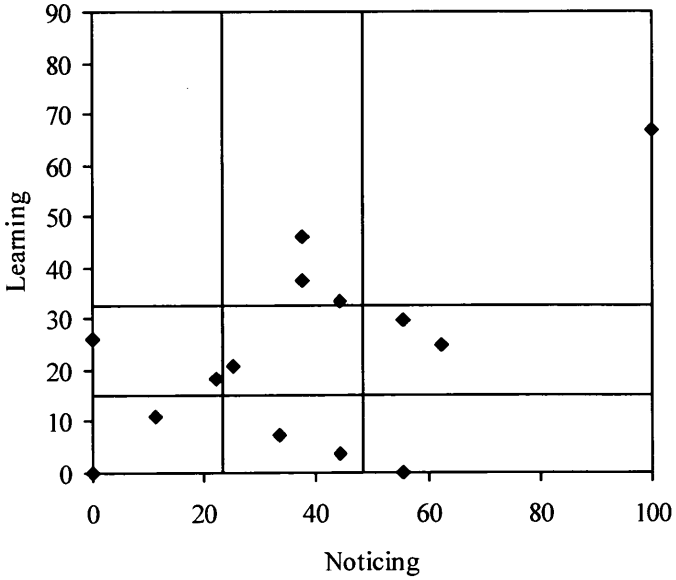


Figure 6.2. Scatter diagram for Round Two vocabulary noticing of target items and vocabulary learning at T1, (r=.56).

[2] Lines in this and following figures indicate points .5 *SD* above and below the mean.

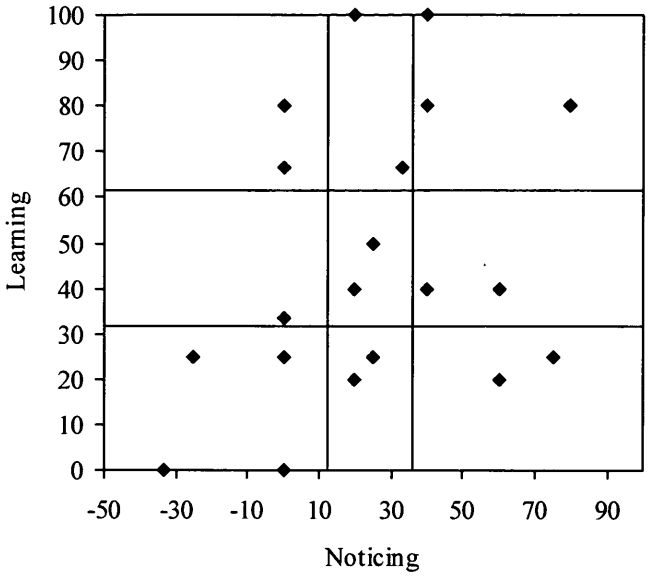


Figure 6.3. Scatter diagram for Round One target grammar noticing and grammar learning shown on the sentence completion task at T1, (r=.35).

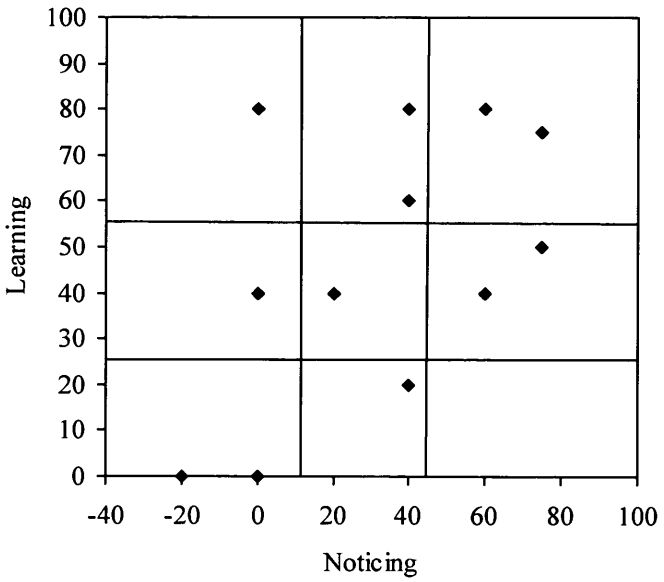


Figure 6.4. Scatter diagram for Round Two target grammar noticing and grammar learning shown on the sentence completion task at T1, (r=.61).

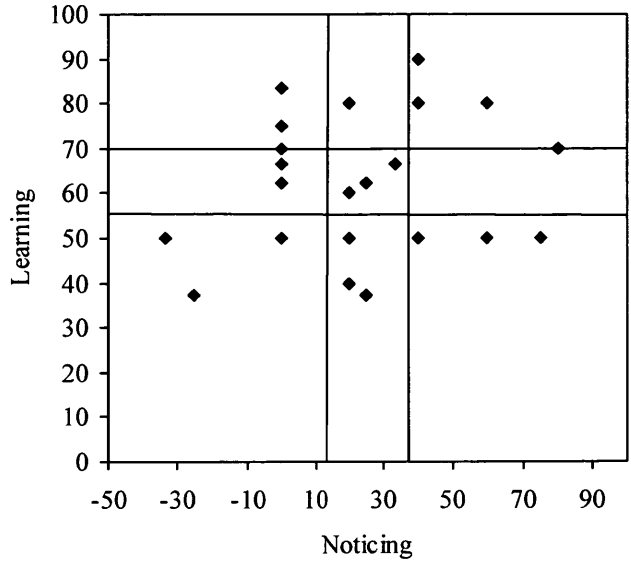


Figure 6.5. Scatter diagram for Round One target grammar noticing and grammar learning shown on the grammaticality judgement task at T1, (r=.22).

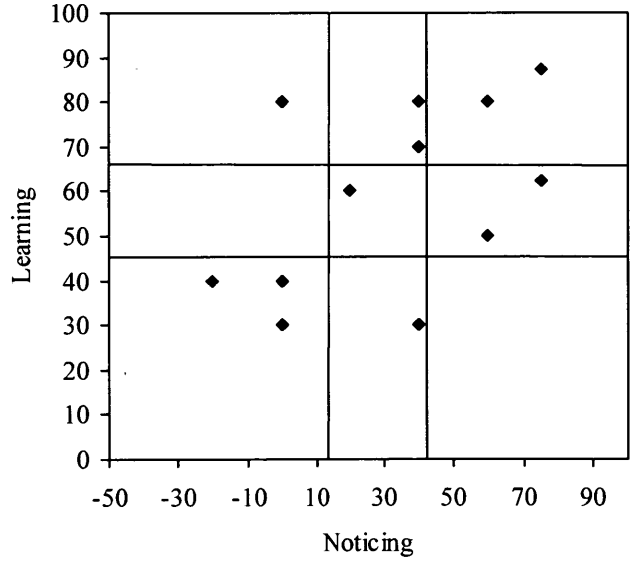


Figure 6.6. Scatter diagram for Round Two target grammar noticing and grammar learning shown on the grammaticality judgement task T2, (r=.56).

And, finally, the same pattern of participants scoring high on learning with and without high scores on noticing can be found in scores for noticing and the grammaticality judgement task, as seen in the following two figures, 6.5 and 6.6.

A summary of all results for participants scoring high on learning is presented in Table 6.5. This table partially summarizes all of the data, showing only those participants who score high on learning. The table lists numbers of individual participants who have fallen into the three categories composed of high learning in combination with low, mid-range, and high noticing. As can be readily seen, the pattern found at T1 of a spread of high learning scores across different levels of noticing is present.

Table 6.5. Summary of the frequency of all high learning scores vis-a-vis low, midrange and high noticing scores using pooled scores.

		Low Noticing	Mid-range Noticing	High Noticing
Grammar (sentence completion)	Round One n=26	T1: 2 T2: 1 T3: 2	T1: 2 T2: 3 T3: 2	T1: 5 T2: 2 T3: 3
	Round Two n=14	T1: 1 T2: 2 T3: 3	T1: 2 T2: 2 T3: 1	T1: 2 T2: 1 T3: 2
Grammar (grammaticality judgements)	Round One n=26	T1: 1 T2: 1 T3: 2	T1: 1 T2: 8 T3: 1	T1: 2 T2: 1 T3: 2
	Round Two n=14	T1: 1 T2: 1 T3: 2	T1: 1 T2: 2 T3: 2	T1: 2 T2: 2 T3: 2
Vocabulary	Round One n=26	T1: 1 T2: 1 T3: 2	T1: 2 T2: 3 T3: 0	T1: 4 T2: 3 T3: 4
	Round Two n=14	T1: 0 T2: 1 T3: 2	T1: 3 T2: 1 T3: 0	T1: 1 T2: 3 T3: 1

6.3. DISCUSSION OF HYPOTHESES 6 AND 7

Hypothesis 6 predicts that there will not be high grammar learning scores without corresponding high noticing scores for grammar. Results relating to this hypothesis were presented in Table 6.5. Counter to the prediction, participants generally ranging in numbers from 1 to 3 (but also including 8 participants in one cell) scored in the High Learning, Low Noticing range or in the High Learning, Mid-range noticing category across both tests of grammar and across all testing times and both data collection rounds. This hypothesis is, therefore, not supported.

Furthermore, there appear to be only slight differences between the measures of grammar learning except for an unusually high number of scores, eight, in the mid-range noticing category for grammaticality judgements at T2 in Round One. This single unusual difference is probably related to the fact that at this testing time, as noted above, there was a convergence of scores between the exposure and instruction groups and that grammaticality judgements may have been a comparatively easier measure on which a higher score could be achieved, as compared with the production task. Furthermore, at this point there was also an unusually low number of participants scoring in the High Noticing, High Learning range, suggesting a shift in the distribution of scores downward as compared to other testing times.

A very similar pattern is found in the data concerning Hypothesis 7, which predicts that no participants will score high on learning vocabulary who do not also score high on noticing vocabulary. As can be seen from Table 6.5, at each testing time across both rounds of data collection there are participants who scored in the Low Noticing or Mid-range Noticing categories who also scored in the High range for learning. Thus, Hypothesis 7 is similarly not supported.

Table 6.6. Summary of the frequency of all the high learning scores vis-a-vis low, mid-range and high noticing scores calculated within the instructed groups.

		Low Noticing	Mid-range Noticing	High Noticing
Grammar (sentence completion)	Round One n=26	T1: 2 T2: 0 T3: 2	T1: 4 T2: 3 T3: 1	T1: 1 T2: 1 T3: 2
	Round Two n=14	T1: 1 T2: 1 T3: 1	T1: 1 T2: 1 T3: 0	T1: 2 T2: 0 T3: 2
Grammar (grammaticality judgements)	Round One n=26	T1: 2 T2: 0 T3: 1	T1: 3 T2: 3 T3: 1	T1: 1 T2: 0 T3: 2
	Round Two n=14	T1: 1 T2: 0 T3: 0	T1: 1 T2: 1 T3: 1	T1: 2 T2: 2 T3: 0
Vocabulary	Round One n=26	T1: 1 T2: 0 T3: 0	T1: 0 T2: 2 T3: 0	T1: 2 T2: 1 T3: 2
	Round Two n=14	T1: 0 T2: 0 T3: 0	T1: 1 T2: 0 T3: 0	T1: 1 T2: 1 T3: 1

Finally, given the results for the similar pattern of scores mentioned above, there appear to be no additional subtle differences in the patterns for the data found between grammar and

vocabulary, thus suggesting a rather strong overall conclusion that there are individuals who can achieve high scores on learning without correspondingly having high scores on noticing.

It may be possible, though, that the conclusions reached above may apply to either the instruction group or the exposure group, but not both groups. The high and low noticing and learning score calculations are based on the pooled data from both groups, as was presented in Section 6.2, where correlations between noticing and learning were initially presented. Thus, a different way of looking at the data in this present context is to provide separate analyses for the instruction and exposure group, justified by the fact that the combined mean scores for noticing and learning lowered the cutoff point for high learning and noticing to the point where the instructed group might easily achieve this level, and perhaps raised the level where the exposure group would have difficulty reaching the "high" level. Again, for both the exposure and instruction groups the cutoff point for "high" and "low" was set at .5 *SD* above and below the mean, calculated separately for each group. This separate analysis is what has been done in Tables 6.6 and 6.7.

Table 6.7. Summary of the frequency of the high learning scores vis-a-vis low, midrange and high noticing scores calculated within the exposure groups.

		Low Noticing	Mid-range Noticing	High Noticing
Grammar (sentence completion)	Round One n=26	T1: 0 T2: 3 T3: 2	T1: 2 T2: 2 T3: 1	T1: 1 T2: 1 T3: 2
	Round Two n=14	T1: 0 T2: 0 T3: 1	T1: 1 T2: 0 T3: 0	T1: 1 T2: 1 T3: 0
Grammar (grammaticality judgements)	Round One n=26	T1: 0 T2: 1 T3: 0	T1: 5 T2: 3 T3: 1	T1: 1 T2: 0 T3: 0
	Round Two n=14	T1: 0 T2: 0 T3: 0	T1: 0 T2: 2 T3: 1	T1: 1 T2: 1 T3: 0
Vocabulary	Round One n=26	T1: 1 T2: 1 T3: 2	T1: 2 T2: 1 T3: 0	T1: 1 T2: 2 T3: 2
	Round Two n=14	T1: 1 T2: 0 T3: 0	T1: 1 T2: 1 T3: 1	T1: 1 T2: 1 T3: 1

Table 6.6 shows the data for the instructed groups, and it is immediately apparent that a similar pattern exists as was found in Table 6.5, where both groups are combined. There are exceptions to the prediction that high learning scores would correspond only with high

noticing scores. Table 6.7 presents data for the exposure group, where a very similar pattern in the data is found. This way of looking at the data, too, provides evidence against Hypotheses 6 and 7.

6.4. FURTHER INTERPRETATION OF NOTICING AND LEARNING DATA

Although Hypotheses 6 and 7 are unconfirmed, there are two points which can be made, which, to a degree, mitigate these results. First, I will look at the pattern of individual scores and look at the frequency of participants scoring in the Low Noticing, High Learning category vis-a-vis the other possible categories. Second, I will return to the correlation data presented above concerning noticing and learning and will discuss the implications of the possible relationship shown there.

For both discussions concerning the general relationships between noticing and learning I will be returning to the original approach to the data represented in the scatter diagrams presented in Section 6.2 and in Table 6.7, where data for both groups are considered simultaneously. I consider it appropriate to use this approach of pooling the data as the cutoff points set for high learning and noticing allow a convenient way of discussing both groups of data simultaneously and establish levels which do indeed intuitively reflect what could be considered high levels of learning and noticing by which both groups can be judged.

6.4.1. ANALYSIS AND DISCUSSION OF HIGH AND LOW LEARNING AND NOTICING SCORES

For this analysis for the frequency of participants scoring in various ranges concerning noticing and learning, data will be summarized and presented together across all three testing times for each round. The reason for doing a summary across all testing times is to look for general trends which may not be apparent at individual testing times, since, as has been seen above, there is variation in both learning and noticing scores across the testing times. Also, it is important to note that this summary is being done with frequency data which have been calculated separately for each testing time, based on the distribution of scores at that time. That is, rather than summarizing actual scores and figuring the frequency that participants fall in to the various categories, frequency data were calculated individually at each testing time. Also, it should be noted that these results are being presented merely descriptively, as statistical testing is not possible due to the fact that scores represented in each cell are not independent of each other. Therefore, these results must be taken rather tentatively. Furthermore, the data that I will be focusing on will be the scores in the corners of the matrixes, that is, the extreme scores in the categories of High Learning, Low Noticing; High Learning, High Noticing; Low Learning, Low Noticing; and Low Learning, High Noticing categories. Thus, any scores falling in the mid-range category for either noticing or learning will only be tangentially referred to. Only those scores will be focused on which were in either in the high or low categories for both noticing and learning.

Data concerning frequency of scores for grammar on the sentence completion task for Rounds One and Two are presented in Table 6.8. As the scores in the corners of the matrices will be the focus of this discussion, they appear in the tables in boldface. In Round One, the basic pattern is present which can generally be found across all the data. Considering the data in the four corners of the matrix, the lowest scores are found for the High Learning, Low Noticing and Low Learning, High Noticing categories, while the frequencies for the High Learning, High Noticing and Low Learning, Low Noticing categories are greater. Concerning Round Two data for grammar on the sentence completion task, the situation is roughly similar, though the frequency for High Learning, Low Noticing is proportionally higher than in Round One.

Table 6.8. Frequency of participants' scores plotting target item noticing and grammar learning of (sentence completion) summarized across all three testing times for Round One and Round Two.

Round One		Noticing		
		High	Mid	Low
	High	**10**	7	**5**
Learning	Mid	9	11	9
	Low	**6**	10	**11**
Round Two		Noticing		
		High	Mid	Low
	High	**6**	5	**6**
Learning	Mid	5	3	3
	Low	**1**	2	**10**

Table 6.9. Frequency of participants' scores plotting target item noticing and grammar learning (grammaticality judgements) summarized across all three testing times for Round One and Round Two.

Round One		Noticing		
		High	Mid	Low
	High	**11**	11	**4**
Learning	Mid	6	9	8
	Low	**8**	8	**13**
Round Two		Noticing		
		High	Mid	Low
	High	**6**	6	**4**
Learning	Mid	4	2	5
	Low	**2**	3	**10**

Similar, though more consistent results are found for frequencies in the grammaticality judgement task, as seen in Table 6.9. Here, in both Rounds One and Two, the High Learning,

High Noticing and Low Learning, Low Noticing categories are both larger than the High Learning, Low Noticing and Low Learning, High noticing categories. This same pattern is then found for vocabulary in Table 6.10.

There are two important trends that then emerge from these data. The frequencies for the High Learning, Low Noticing category are lower than the High Learning, High Noticing category in five out of six cases. That is, the more extreme exceptions to Hypotheses 6 and 7 are actually a relatively small number when compared to the predicted outcome. Furthermore, in all cases shown above, the frequency for participants who fall into the Low Learning, High Noticing category is consistently lower than the Low Learning, Low Noticing category. That is, there are relatively few cases of people who are bad learners and good noticers. This means that in looking at these more extreme cases, high learning is associated with high noticing, while low learning is associated with low noticing. These results, though, do need to be treated rather cautiously as the numbers are small, and further statistical analysis cannot be done on them. Nevertheless, the above discussion can serve to mitigate the rather strong rejection of Hypotheses 6 and 7 and suggest a relationship between noticing and learning.

Table 6.10. Frequency of participants' scores plotting target item noticing and vocabulary learning summarized across all three testing times for Round One and Round Two.

Round One		Noticing		
		High	Mid	Low
Learning	High	11	5	4
	Mid	8	9	10
	Low	5	10	16
Round Two		**Noticing**		
		High	Mid	Low
Learning	High	5	4	3
	Mid	4	6	4
	Low	4	7	5

A final and important point here concerning the interpretation of these results should be made. These results need to be looked at in terms of whether the learning without noticing occurred in the instructed condition or in the exposure condition. For the instructed condition, it would be possible to claim that these participants who have high learning and low noticing scores are those participants who have been instructed to and have learned to manipulate the grammatical forms independently of noticing them. There is indeed evidence in individual variation concerning the ability to notice, as was presented in Chapter 7, and this suggests that some learners may simply be less aware than other learners. Thus, in the instruction group, these learners would simply be learning without showing awareness of target items in input, and the two factors would indeed not be connected as learning would have been induced through instruction.

Those who have shown high learning and low noticing in the exposure group are more difficult to explain. These would appear to be clear cases where learning through input has happened without the benefit of noticing and awareness. Again, the definition of what is "high" needs to be discussed here. If each group is considered individually and a cutoff point for high learning and noticing is established with reference to mean scores, then one runs the risk of validating relatively unimportant levels of noticing and learning as "high" if the mean scores for that group are low. If the results from both the exposure and instruction groups are pooled and mean scores used to define "high", then a stricter assessment of high learning and noticing can be made, as was argued in Chapter 6. If the results for the exposure group alone are extracted from the pooled data represented in Tables 6.8, 6.9 and 6.10, then the following picture emerges. These tables presented data categorizing learners on low, medium and high scores on learning and noticing, with learners falling into the high learning, low or medium noticing categories being exceptions to the idea that noticing and learning should co-occur. But if only the exposure group is looked at, across both data collection rounds and all three testing times, there are only three instances of participants from the exposure group scoring in this range for vocabulary, five for grammaticality judgements, and one case in the sentence completion task. That is, out of all of the tests conducted for grammar and vocabulary across all of the data, there are only nine cases out of the 240 (roughly 4% of the total) which show participants from the exposure group having high learning without high noticing. Looking at this situation this way, these most difficult to explain cases represent a very small number indeed and could simply be treated as exceptions caused either by imperfect pre-testing, good guesses, or exceptional abilities.

6.4.2. FURTHER DISCUSSION OF CORRELATIONAL DATA

Another way of looking at the relationship between noticing and learning — and a way which provides rather stronger data than the above argument — is to look at the correlations between learning and noticing that were presented in Section 6.2. The lack of support for Hypotheses 6 and 7 should not imply a total lack of a relationship between noticing and learning. Indeed, evidence has been presented in Table 6.4 in section 6.2 which shows that, in almost all cases, positive correlations for noticing and learning were found. Furthermore, in many cases these correlations are relatively large correlations, particularly at T1. In fact, this correlational evidence in many cases provides rather solid support for a possible relationship between noticing and learning, though lacking the force of evidence if support for Hypotheses 6 and 7 had been found by finding no exceptions to high learning corresponding to high noticing. It is, of course, impossible to show a causal relationship between noticing and learning, though the correlations do at least suggest a co-occurrence of noticing and learning in many cases.

Second, another issue which can provide additional support for a relationship between noticing and learning concerns the issue of whether participants made Remember or Know responses to items they noticed. So far, in all of the analyses only Remember responses have been addressed, as they have formed the definition of what it is to "notice". Remember responses associated with learning would imply, by definition, a relationship or co-occurrence

of consciousness and learning. On the other hand, Know responses being associated with learning would suggest a lack of conscious memories related to or co-occurring with learning. This is an important issue, as a stronger association of consciousness with learning than lack of consciousness (as defined by Remember or Know responses respectively) would either strengthen or weaken the claim represented in Hypotheses 6 and 7 that learning and noticing are related in some way. That is, a stronger correlation between Remember responses and learning than Know responses and learning would add some supporting evidence of a relationship or co-occurrence between learning and consciously noticing items in input.

Table 6.4 presents data showing the correlations of learning scores both with Remember and Know responses. Excluding for the moment the correlations concerning the grammaticality judgement task, in the overwhelming majority of the cases, Remember responses are associated with stronger — and, in many cases, much stronger — correlations than Know responses. For vocabulary, Remember responses have higher correlations with learning than Know responses in every case. Concerning the sentence completion task, correlations for Remember responses with learning are higher in 5/6 of the cases than Know responses. These data thus support the claim that learning and consciousness are associated to some degree, as Remember responses represent incidences of retrieval of information form episodic memory. The interesting exceptions to this are found in the grammaticality judgement task where in four out of the six cases the correlations between learning and the Know responses are higher than the correlations with the Remember responses with statistically significant results being found in Round Two at T1. Of the three measures of learning, the grammaticality judgement task is the one most likely to involve an intuitive judgement on the part of the participants, and therefore it does not seem surprising that the exceptional correlations with the Know responses would be found there. It is the Know responses which are associated with strong feelings of familiarity and with the absence of conscious memories, and for this reason they have sometimes been considered to be associated with implicit memory (Gardiner & Parkin, 1990). Thus, it makes sense that there are a number of participants whose Know responses, reflecting a lack of conscious knowledge and perhaps implicit memory, would correlate higher with a measure of grammar knowledge that is more likely to tap intuitive, or implicit knowledge.

6.5. RESULTS WHEN USING REMEMBER + KNOW SCORES TO CALCULATE NOTICING

As was raised in the previous chapter, it may theoretically be possible to define noticing as being demonstrated by both Remember and Know responses. Therefore, it is important to look at the relationship between learning and noticing and see if a different picture emerges when noticing is defined in this broader way.

Table 6.11 summarizes the results. Again, caution must be used in interpreting the correlation results. These results support the general trend as seen with the Remember responses seen in Table 6.4., and appear to be more stable and consistent, just as the

Remember + Know data were for noticing above. Furthermore, using the Remember + Know data, there are fewer negative correlations and more cases of significant results. This carries through to the data for the grammaticality judgements which, when the Remember and Know responses are both used, appear to be in line with the other results, unlike then the Know data alone was looked at. Thus in general the Remember + Know responses appear to provide stronger and more consistent data, though the issue still remains of whether a subtle distinction is being lost by combining the two types of responses together.

Table 6.11. Correlations between learning and accuracy of noticing for Rounds One and Two considering Remember and Know responses together.

	T1	T2	T3
Round 1			
Grammar noticing x sentence completion task	.58	.27	.05
p	.05	.187	.821
Grammar noticing x grammaticality judgement task	.35	.22	.01
p	.082	.281	.952
Vocabulary noticing x vocabulary learning	.58	.31	.52
p	.002*	.121	.006*
Round 2			
Grammar noticing x sentence completion task	.74	.35	-.24
p	.002*	.215	.417
Grammar noticing x grammaticality judgement task	.77	.44	.72
p	.001*	.117	.004*
Vocabulary noticing x vocabulary learning	.56	.22	.005
p	.031	.46	.985

6.6. SUMMARY CONCERNING A RELATIONSHIP BETWEEN NOTICING AND LEARNING

In the previous chapter, it was established that noticing did indeed occur, and in many cases it occurred to a greater degree with the instructed group. In this current chapter, first it was established that learning did occur and that in each case learning occurred to a greater degree in the instructed than in the uninstructed group.

Secondly, the issue of Hypotheses 6 and 7 was addressed, concerning the prediction that there would be no cases of participants who scored high on learning but did not score high on noticing. A somewhat arbitrary decision was made to define "high" as occurring .5 *SD* above the mean for the combined data for the instruction and exposure groups. An analysis of these data showed that, contrary to predictions, exceptions were found for vocabulary and grammar at each testing time and across both data collection rounds, and thus the hypotheses were not supported. Furthermore, it was shown that in an analysis of the same issue when the exposure and instruction groups considered separately and +.5 *SD* "high" marks were established for each group, that the conclusions remained the same: exceptions were found at nearly every testing time for both grammar and vocabulary, across both testing rounds.

Third, despite the rejection of Hypotheses 6 and 7, a relationship or at least a co-occurrence between noticing and learning was supported by three lines of argument. First, it was shown that if the more extreme scores are looked at, higher learning is generally associated with higher noticing, and lower learning is associated with lower noticing. Second, in a similar analysis, the correlational data were returned to in order to show that indeed in most all cases a positive correlation between noticing and learning was found, and that in some cases these correlations were rather strong, particularly at T1. Finally, this same issue was addressed from the point of view of the kinds of responses that participants made when indicating that they noticed an item in the noticing text, and it was shown that there were generally much stronger correlations between Remember responses with learning than Know responses with learning, thus making the case that learning is indeed associated, or at least co-occurs, with consciousness, as defined by the provision of a Remember response.

Finally, the extended definition of noticing was used including both Remember and Know responses. The results from this analysis were very similar to the results just using the remember data, except that the trends were strengthened and more cases of statistically significant data were found.

Thus, while an absolute and exceptionless relationship between noticing and learning was clearly not found, there is good evidence to suggest that a relationship does exist at least to some degree and in certain situations.

CHAPTER 7. INDIVIDUAL DIFFERENCES AND INDIVIDUAL VARIATION IN NOTICING

In this section I will look at scores for individual participants for noticing and explore the issue of whether individuals can be characterized consistently across the data as high or low ability for noticing. First, the consistency of individual noticing scores will be explored by looking at correlations for noticing across testing times and between grammar and vocabulary. Second, a general profile of participants who are good and poor noticers will be established and discussed.

This individual analysis is being done for two reasons. First, although the data presented here do not have direct bearing on the acceptance or rejection of the hypotheses of this research project, they can be used to provide indirect support for the claim that instruction has an impact on increasing the noticing of grammar and vocabulary, as will be discussed below. Second, little or no research has been done to date concerning the question of individual differences in noticing, and it is possible that there may be individuals for whom instruction has a large impact on their noticing of grammar and vocabulary, while others may simply be "naturally" good or bad at noticing, for whatever reason. This is an important issue to look at which can fill a gap in the literature, as well as have bearing on the key issues raised in this paper.

7.1. CONSISTENCY OF NOTICING SCORES ACROSS TIME SHOWN BY CORRELATIONS

One way that the above issues can be addressed is by looking into whether or not individuals have consistent noticing scores across different testing times. The data generally show that individuals who are instructed score more consistently across each testing time, and this suggests an influence of instruction in making responses more consistent and homogenized, while responses of uninstructed learners can be characterized as more random and attributed to different individual participants' noticing at each particular testing time.

Table 7.1 presents correlation matrices for grammar noticing in Round One and Round Two which show correlations of results from T1 with T2 and T1 with T3. Round One data show that the correlations for the instructed group are both larger and more consistent than

those for the exposure group. While the correlations at T2 are similar between the two groups, at T3 the exposure group shows a rather large negative correlation suggesting a shuffling of the consistency of participants across testing times. Round Two results, though, show roughly the opposite picture, with stronger correlations in each case. Here it is the exposure group which has more consistent scores than the instruction group, with the instruction group showing a rather large negative correlation between noticing scores at T1 and T3. These large negative correlations are rather difficult to explain and must be seen to underscore the fact these data are rather unreliable and must be treated with caution. This is especially the case with the correlations in Round Two where groups were comprised of only 7 participants.

Table 7.1. Correlation matrices for grammar noticing for Round One and Round Two showing correlations of T1 with T2 and T3 for target items.

	T2	T3
Instruction group at T1		
Round One	0.213	0.201
Round Two	0.496	−.698
Exposure group at T1		
Round One	0.159	−.514
Round Two	0.86	0.413

The results for vocabulary, shown in Table 7.2, present a more consistent picture than the results presented above for grammar. In both Round One and Round Two, the instruction group shows both larger correlations than the exposure group and also more consistent results. Furthermore, correlations for the instruction group are quite large in some cases, such as the .80 and .96 correlations at T2 for Round One and Two, respectively.

Table 7.2. Correlation matrices for vocabulary noticing for Round One and Round Two showing correlations of T1 with T2 and T3 for target items.

	T2	T3
Instruction group at T1		
Round One	0.795	0.702
Round Two	0.963	0.697
Exposure group at T1		
Round One	0.226	0.043
Round Two	0.413	0.217

Thus, the data in Tables 7.2 provide good evidence, at least in the case of vocabulary, that instruction has an effect in allowing for partial consistency of scores across testing times. This consistency, where found, most likely results from the fact that instruction provides added saliency for those participants who are already likely, due to whatever individual characteristics, to notice something, whereas for the exposure group, at each testing time the results

are primarily influenced by the features that make the grammar or vocabulary items salient in the texts as well as participants' personal tendencies towards noticing. This can, then, provide supporting evidence for the position that instruction has a positive effect on noticing, at least in the case of vocabulary, while the results for grammar are mixed.

7.2. CONSISTENCY OF SCORES SHOWN BY INDIVIDUALS

Another way of looking at the consistency of scores is to identify individuals who score consistently well or poorly across all testing times. Evidence of this, while supporting the above conclusions concerning the effects of instruction could also be used to support a view whereby the ability to notice elements in input is partially influenced by a learner's natural tendencies to do so. The issue here is to identify these participants and look in more detail at their profile as noticers and learners.

Three things need to be noted concerning the calculations for scores concerning individual performance. First, for this section and the following section, a definition of a "good" or "bad" noticer will be defined similarly as it was in section 6.2 above, concerning the relationship between learning. Thus a cutoff point of +/- .5 *SD* above or below the mean will be used to define good and bad noticers. Here, a separate mean was established for each group, which was different than the analysis that was done for learning and noticing, where the results of the two groups were pooled together. The reason that this was done was to establish the definition of high or low noticing within the contexts of the treatment that the different groups underwent. Since noticing scores for exposure were generally lower than those for instruction, a definition of high and low noticing based on individual groups does not bias results in favor of the instruction group. Thus, for example, a "good" noticer could achieve a top score in the instructed condition and a top score in the exposure condition, and the two scores could be quite different from each other. Calculating separate means for each group allows the possibility of capturing the fact that each score is a high score within its own context.

Second, it is important to note that in order to have a general picture of how individual participants performed concerning their ability to notice, scores were again summarized across all three testing times. While variation in noticing scores has been noted above — particularly concerning the exposure group — the advantage of having the ability to make generalizations for participants considering all of the data together outweighs the disadvantages of losing the ability to describe subtle changes in performance over time.

Finally, again it should be pointed out that participants served simultaneously as the instructed group for one target area, and the exposure group for the other. Thus, it is possible to plot individual participants' noticing scores for both instruction and exposure, as has been done in the following figures.

Overall, there are a number of points which can be made concerning individual differences in noticing. Results for Round One are shown in Figures 7.1 and 7.2 and establish the basic pattern which is also found in Figures 7.3. and 7.4 for Round Two. First, it should be noted

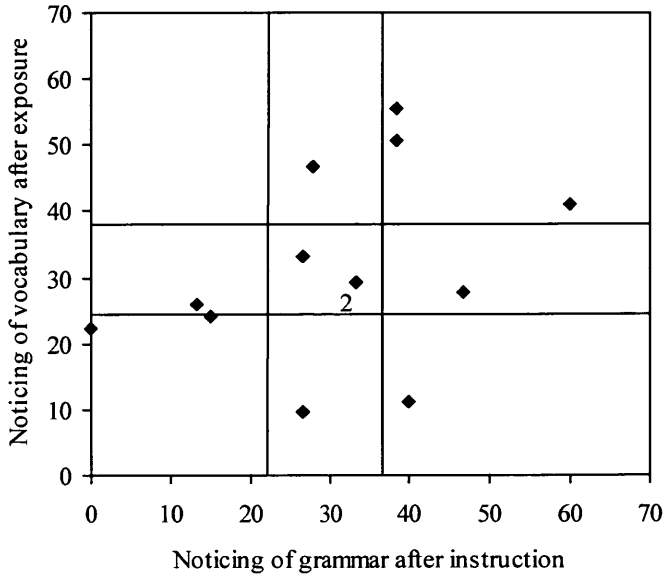


Figure 7.1. Target item noticing scores for Round One grammar instruction/vocabulary exposure group summarized across all three testing times (r=.34). (Note: "2" = two participants received the same score.)

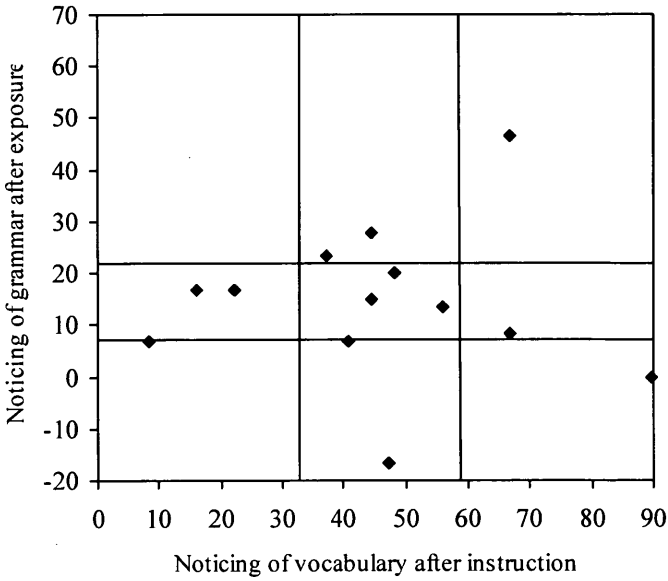


Figure 7.2. Target item noticing scores for Round One, vocabulary instruction/grammar exposure group summarized across all three testing times (r = –.004). (Note: "2" = two participants received the same score.)

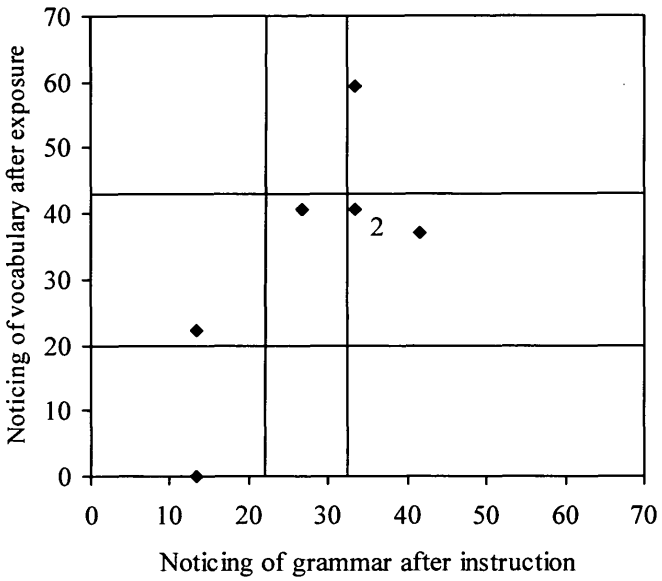


Figure 7.3. Target item noticing scores for Round Two, grammar instruction/vocabulary exposure group summarized across all three testing times (r = .75). (Note: "2" = two participants received the same score.)

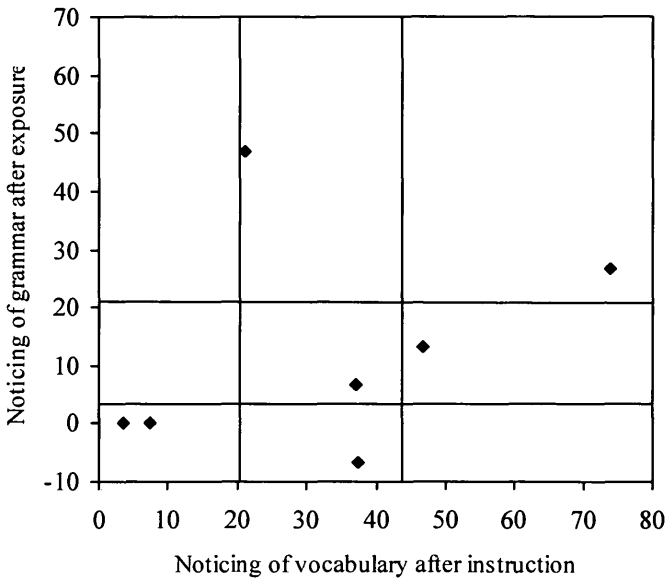


Figure 7.4. Target item noticing scores for Round Two vocabulary instruction/grammar exposure summarized across all three testing times (r = .28).

that participants who score in the High, High or Low, Low range can be found. There are four participants in the Round One and two participants in Round Two who score in the high range for both instruction and exposure. Thus, these participants respond well to the effects of instruction encouraging their noticing and also appear to have heightened awareness of vocabulary or grammar items in input in general. In the Low, Low range there were three participants in Round One and three participants in Round Two. For these participants, instruction clearly has no large effect on their ability to notice target items, and they appear largely unaware of items which they are exposed to.

Second, the opposite two categories, high instruction, low exposure noticing and low instruction, high exposure noticing are largely or completely unrepresented in the data. Across both rounds, no scores were found in the low instruction, high exposure range. While this might very well be expected in the case of participants who received instruction in the more surface level of vocabulary and only exposure to the more abstract level of grammar, the opposite case holds true as well, where those instructed in grammar do not score high on vocabulary exposure unless their grammar noticing scores are in the high or medium range. Concerning the high instruction, low exposure range, only two participants scored within this category, both in Round One. These participants are thus those whose noticing is heavily tied to instruction in that they appear to have only quite limited ability to notice items which they are merely exposed to. That few participants score within the two categories described here lends more evidence to the possibility that participants' natural ability to notice interacts with instruction.

Third, when these participants in the three extreme groups represented in the data are compared across the other variables of learning and proficiency, a somewhat consistent pattern emerges, though exceptions are present which prevent any strong general conclusions from being made. Tables 7.3 and 7.4 present data on learning and proficiency for the participants for each round who scored in the extreme categories. They are listed by participant number and divided into two groups with the grammar instruction/vocabulary exposure group represented in Table 7.3, and the vocabulary instruction/grammar exposure group in Table 7.4. Figures in these tables are given in percentile scores figured for each group in order to allow for the general comparison across participants within the same group. Looking at Table 7.3, it can be seen that participants who scored high on noticing for the instructed items and also for exposure items (the "High High" group) tended to score well on other variables as well. Participant 16 could be considered the model case here, as this person scores well across all categories for learning and also general proficiency, which was tested before the study began. This participant's scores are also impressive for having scored so well on vocabulary learning, as this group received no instruction on vocabulary items. Participants 21 and 29 also stand out as what appear to be exceptions when compared to the expected results found for participant 16. Scores for participant 21 are just slightly above the 50th percentile or below, while all of participant 29's scores are at or below the 50th percentile. While participant 16's high noticing corresponds with high learning and high general proficiency, participant 29's scores do not. Looking at Round Two data on the same table, participant 33, again, is a participant who scores in the higher ranges on many of the tests, particularly gram-

mar learning, which was the instructed condition for this group, though scores for vocabulary learning and grammatical proficiency are just below the 50th percentile. A similar pattern for the High, High group can be found in Table 7.4, where the results for the vocabulary instruction/grammar exposure groups are presented across both rounds. Here, again, the two participants in the High, High range, participant number 31 in Round One and number 29 in Round Two, show a mixture of high, mid-range, and sometimes quite low scores. The conclusions that can be drawn from both tables is that the expected results of consistently high scores across all measures for those in the High, High groups are actually unfounded. The only participant who achieves these results is participant 29 in Round One. The other participants do have high scores, particularly in at least one area of learning, but have a mixture of high mid-range and low scores across all tests.

Table 7.3. Learning and proficiency percentile scores for participants from the grammar instruction/vocabulary exposure groups who scored in the three extreme categories for noticing target items.

Participants falling in each category	Vocab-ulary learning	Grammar learning (sent. complet.)	Grammar learning (gram. judge.)	Vocab-ulary pro-ficiency	Grammar pro-ficiency
Round 1					
High instructed noticing/ High exposure noticing					
#16	91	83	75	83	83
#21	50	8	58	25	58
#29	1	50	41	33	16
Low instructed noticing/ Low exposure noticing					
#34	1	25	66	75	1
#28	1	33	16	41	33
High instructed noticing/ Low exposure noticing					
#32	1	1	75	8	8
Round 2					
High instructed noticing/ High exposure noticing					
#33	46	92	84	72	44
Low instructed noticing/ Low exposure noticing					
#42	1	16	66	16	66
High instructed noticing/ Low exposure noticing (no participants)					

Table 7.4. Learning and proficiency percentile scores for participants from the vocabulary instruction/grammar exposure groups who scored in the three extreme categories for noticing target items.

Participants falling in each category	Vocab-ulary learning	Grammar learning (sent. complet.)	Grammar learning (gram. judge.)	Vocab-ulary pro-ficiency	Grammar pro-ficiency
Round 1					
High instructed noticing/ High exposure noticing					
#31	75	8	1	41	91
Low instructed noticing/ Low exposure noticing					
#42	16	8	83	25	66
High instructed noticing/ Low exposure noticing					
#25	99	99	58	99	99
Round 2					
High instructed noticing/ High exposure noticing					
#29	99	50	1	50	33
Low instructed noticing/ Low exposure noticing					
#32	1	1	33	16	16
#37	16	16	50	1	66
High instructed noticing/ Low exposure noticing					
(no participants)					

For the participants who scored low on noticing for both the instructed and uninstructed conditions, the "Low Low" group, the picture is somewhat clearer and the generalization can be made that these participants score particularly low on learning and have somewhat lower scores in proficiency. Looking at Table 7.3, all of the Low, Low participants, numbers 34 and 28 in Round One and 42 in Round Two, score quite low or the lowest possible in learning, except for grammaticality judgements, where scores in general tended to be higher for all groups and tended to be clustered together more tightly across a narrower range of scores than were found for the combined vocabulary learning and sentence completion grammar scores. For the instructed group, the mean for the combined scores for grammaticality judgements was 67.63 with a standard deviation of 9.05, and for the exposure group the mean was 57.35 with a standard deviation of 8.78, and in both tests the range between the high and low scores was smaller than on other tests. Thus the difference between a high and low percentile ranking can be a matter of a few points on this test. It is possible, then, to discount a high score on grammaticality judgements if scores on other measures were low. A similar pattern

for Low, Low scores is presented in Table 7.4. Thus, a general conclusion here is that consistently low scores on noticing correspond with consistently low scores in learning and generally, but not always, lower scores on the two measures of proficiency.

Across all of the data, there were only two participants who scored in the high noticing for the instructed condition and low noticing for the uninstructed condition, and these participants present a series of rather exceptional and contradictory scores which preclude any speculative conclusions. In Table 7.3, participant 32 scores in the lowest percentile for learning, while in Table 7.4, participant 25 scores in the highest percentile for both tests of learning. Given that participant 25 also scored in the highest percentile for both vocabulary and grammar proficiency, while participant 32 scored in the lowest percentile for those two measures of proficiency, it is rather difficult to speculate as to why these two participants ended up scoring in the same "High Low" category for noticing.

The general pattern, then, that emerges concerning the relationship between noticing and learning and proficiency, is that those who score in the "High High" range for noticing tend to have high, but not always the highest scores for learning and proficiency, while those who score in the "Low Low" range tend to have consistently low scores across the learning and proficiency. At least looking at these extreme cases, then, this suggests that noticing is related to the other variables of learning and proficiency.

A fourth point which can be made about individual differences in noticing is that individuals remain rather consistent in their scores on noticing. This can be seen by looking at the subgroup of participants from Round One who also participated in Round Two. Figure 7.5 shows a comparison of these participants on scores for the noticing of target items in each round, with scores from each round being represented together on the figure. Circles show participants who remained in the same range of scores for noticing in both rounds, while arrows show changes from Round One to Round Two. Of the 13 participants who took part in both rounds, three remained in the same category. Five participants moved up or down on a single dimension. For example, participant 23 moved from Medium instruction, High exposure noticing to Medium instruction, Medium exposure noticing. Three participants changed both categories, though each only on a single dimension. For example, participant 27 moved from "Medium High", to "High Medium". Finally, only two participants, numbers 38 and 32 moved more than two categories between the two rounds.

Thus, overall, results are rather consistent across the two rounds. This conclusion is important for two reasons. First, it does suggest that a certain amount of reliability of results concerning the noticing test itself can be reflected in these data. That is, individuals repeatedly tested do produce similar results on the test. Second, these results suggest a certain consistency of noticing across grammar and vocabulary due to the fact that those participants who were in the grammar instruction/vocabulary exposure group in Round One were put in the vocabulary instruction/grammar exposure group in Round Two and vice versa. Thus, where these results are consistent, they show consistency across both grammar and vocabulary. This points to either the consistent influences of individual differences in noticing, or the consistent effect of instruction (or lack of effect of instruction in some individual cases) or a combination of these factors, although it is impossible to decide which factors are producing this relative consistency.

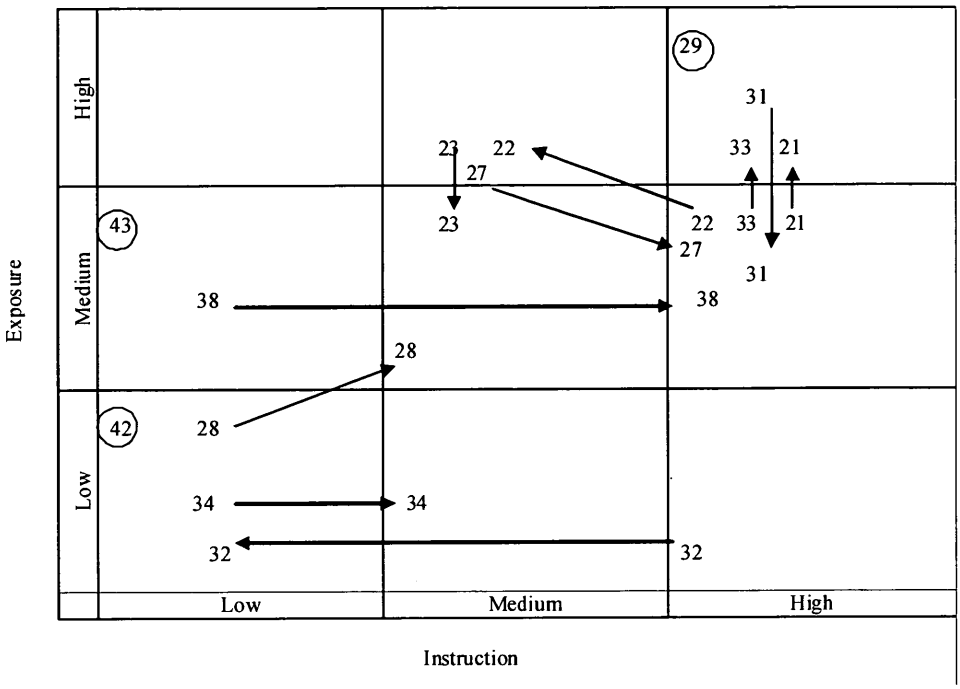


Figure 7.5. Summary of target item scores for participants who participed in both Round One and Two, showing differences between Round One and Round Two scores. (Circles indicate no change in category, while arrows do.)

A fifth point which can be made by looking at these individual scores is to reinforce what was said in earlier sections concerning the relative ease of vocabulary noticing as compared to grammar noticing. Looking back at Figures 7.1 through 7.4, which plot noticing for instruction and exposure, correlations can be calculated to the degree to which noticing of the instructed items co-occurs with the noticing of the items which were the object of exposure. Correlations calculated for Figures 7.1 and 7.3, that is, between grammar instruction and vocabulary exposure are +.34 and +.75, respectively, for Round One and Round Two, neither being statistically significant. Correlations calculated for Figures 7.2 and 7.4, between vocabulary instruction and grammar exposure are -.004 and +.28 respectively, for Round One and Round Two, again, without being significant. Thus, ignoring the issue of statistical significance, there is a higher correlation between the noticing of instructed grammar and the noticing of vocabulary which was the object of exposure, than for grammar which was the object of exposure and vocabulary instruction. A possible explanation is that when grammar is presented in the exposure condition, it appears more difficult to notice than when vocabulary is presented in the exposure condition, otherwise these two sets of correlations would be expected to be more similar. Again, though, the results of correlations between such small groups of participants must be interpreted with caution.

7.3. SUMMARY AND CONCLUSIONS

The evidence presented in this section on individual differences leads to several tentative conclusions. First, it was established that results in general were relatively stable across time. Second, it was established looking across all three testing times in both data collection rounds that groups of learners could be identified who could be characterized as high or low noticers, that these tendencies seem, in the case of high noticers, to co-occur with high scores in learning and on proficiency tests, and that these characteristics of being a high-level or low-level noticer potentially remain consistent across time. Finally, further evidence was given supporting the claim that vocabulary is easier to notice than grammar in the exposure condition.

CHAPTER 8. THE NOTICING OF INDIVIDUAL TARGET GRAMMAR AND VOCABULARY ITEMS

In this chapter, results will be presented concerning the noticing of individual target grammar and vocabulary items. Four different issues will be raised here. First, a general overview will be presented comparing the instructed and exposure groups over time. Next, individual items will be analyzed to get some perspective on which items were most susceptible to the effects of instruction. Third, items will be discussed in terms of which ones were most easily noticed. Finally, the issue of differences between the instruction and exposure group will be raised in terms of each group's order of accuracy concerning the noticing of particular items. It is important to note here that throughout this section, the data on individual items are expressed in terms of an accuracy score just in the same way as accuracy scores for individual participants were calculated for the analysis of noticing presented earlier. Thus, negative scores will be found indicating that an item received more incorrect Remember responses than correct Remember responses for noticing by participants.

8.1. OVERVIEW OF RESULTS ON INDIVIDUAL ITEMS

In comparing the instructed group and the exposure group on individual items, the overall picture, of course, reflects what was presented above concerning the comparison between the experimental groups on mean scores for noticing, where a strong performance for the instructed group at T1 is followed by weaker performance at T2 and T3. Looking at the individual results, two points can be made. The first one reinforces the above results and the second adds a new perspective by looking at the results across time. It should be noted that due to the multiple comparisons that would need to be made for the data on individual items, testing for statistical significance of comparisons cannot been done.

First, it can be seen across both the grammar and vocabulary scores for individual items that the instructed group performs better on almost every individual item at T1, while results are mixed at T2 and T3. This suggests that an effect of instruction is to heighten the noticing of all target items regardless of type and regardless of any factors relating to their position in the texts for noticing. Yet, this effect of instruction is largely only present in the short term.

For grammar in Table 8.1 it can be seen that in Round One at T1, the instruction group outperforms the exposure group on each of the five target items except for the "*have* a verb"

construction, where the exposure group has a higher score. The difference on this item between the two groups is the smallest found for this round, and is the result of a single participant in the exposure group reporting the noticing of the construction. In Round Two, Table 8.2, again the instructed group scores better than the exposure group on each construction at T1.

Table 8.1. Noticing scores for individual target grammar items for Round One for the grammar instruction and grammar exposure groups.

	T1		T2		T3	
	Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
Cleft sent. w/ subject focus	30.77	–7.69	7.69	15.38	23.08	7.69
Pseudo-cleft emph. verbs	46.15	–23.08	53.85	30.77	7.69	–7.69
"give a verb"	92.31	76.92	69.23	61.54	61.54	53.85
possessive + gerund	23.08	0	–7.69	7.69	—	—
"have a verb"	0	7.69	—	—	15.38	–7.69
"get" passive	—	—	–7.69	–15.38	15.38	7.69
Note: Dashes indicate that an item was not tested at that time.						

Looking at the pattern for grammar at T2 and T3 across both rounds, the situation is generally different. For Round One, results at T2 are mixed, though at T3 the instructed group again outperforms the exposure group, this time on each item. In Round Two at T2 the results are mixed, with the instruction group performing better in two cases, yet having the exact same results at the exposure group in three cases. In Round Two at T3, the results are again mixed, with the exposure group doing better in three out of the five cases. Thus, with the exception of the scores at T3 in Round One, the advantage for instruction can be said to generally fall off across the testing times.

The results for individual vocabulary items are presented in Tables 8.3 and 8.4. Looking at the results at T1, again the pattern can be found as described above, where the instructed group does consistently better than the exposure group. This is clearly the case in Round One, where the instructed group scores better on each item except the word *sustain*, where results for the two groups are equal. For Round Two, the results are not as strong, though the pattern is still present. The instructed group scores better on six out of the nine items, with the exposure group scoring better on two, and equal results between the two groups found on one item.

Table 8.2. Noticing scores for individual target grammar items for Round Two for the grammar instruction and grammar exposure groups.

	T1		T2		T3	
	Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
Pseudo-cleft *is...is*	71.43	28.57	14.29	14.29	57.14	42.86
despite + gerund	57.14	28.57	28.57	–14.29	28.57	0
recommend + gerund	14.29	–14.29	0	0	0	14.29
conditional + inversion	28.57	14.29	42.86	14.29	—	—
adverb + inversion	42.86	0	—	—	0	14.29
Judgment verb + *be*	—	—	14.29	14.29	0	28.57
Note: Dashes indicate that an item was not tested at that time.						

Table 8.3. Noticing scores for individual target vocabulary items for Round One for the grammar instruction and grammar exposure groups

	T1		T2		T3	
	Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
deliberately	61.54	23.08	30.77	46.15	69.23	30.77
expose	53.85	0	38.46	23.08	38.46	30.77
principle	53.85	46.15	61.54	69.23	76.92	61.54
betray	46.15	30.77	—	—	30.77	30.77
condense	—	—	15.38	15.38	15.38	15.38
depict	—	—	30.77	7.69	30.77	23.08
deteriorate	53.85	38.46	61.54	76.92	—	—
profound	69.23	46.15	23.08	61.54	—	—
propensity	38.46	7.69	—	—	53.85	23.08
succession	69.23	53.85	7.69	7.69	—	—
sustain	46.15	46.15	—	—	15.38	15.38
trigger	—	—	69.23	76.92	61.54	15.38
Note: Dashes indicate that an item was not tested at that time.						

Looking at vocabulary across testing at T2 and T3, results are, again, mixed. This is the case clearly at T2 in both rounds. In Round One at T2, the exposure group has better results than the instruction group on five of the nine items, while in Round Two, both groups have high scores on three items and then have equal results on the remaining three items. At T3 in Round One, the instruction group, as was the case for grammar in Round One, recovers to some extent and shows much more consistent results than at T2. In six of the nine cases, here the instructed group does better, and in the remaining three cases, results are equal for both groups. In Round Two at T3 the results are, again, mixed, with the exposure group performing better on five of the nine items, the instructed group doing better on two items, and equal results between the two groups on two items.

Table 8.4. Noticing scores for individual target vocabulary items for Round Two for grammar instruction and grammar exposure groups

	T1		T2		T3	
	Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
bulk	71.43	42.86	28.57	57.14	42.86	85.71
compel	28.57	14.29	28.57	28.57	0	14.29
tangible	42.86	57.14	57.14	42.86	57.14	85.71
ambiguous	57.14	42.86	42.86	28.57	—	—
consent	42.86	14.29	—	—	0	0
contemplate	28.57	14.29	14.29	0		
cumbersome	—	—	28.57	57.14	42.86	42.86
enhance	—	—	28.57	42.86	14.29	28.57
evoke	28.57	42.86	42.86	42.86	—	—
flurry	85.71	71.43	—	—	57.14	28.57
incentive	—	—	14.29	14.29	–14.29	0
intrinsically	14.29	14.29	—	—	42.86	28.57

Note: Dashes indicate that an item was not tested at that time.

Thus, looking at grammar and vocabulary across both rounds, there is a general picture of consistent superior results on most all items for the instructed group at T1, followed by mixed results at later testing times. The exceptions to this occur in both grammar and vocabulary in Round One, where the instructed group recovers to have consistently better results on every item at T3. Thus, this points to the overall effect that instruction has on increasing the noticing of all target items at least in the short term.

A second general point that can be made concerns comparisons of scores on individual items across all three testing times and deals with the issue of which group, the instructed group or the exposure group, shows consistently better results on individual items across

time. The results presented in Tables 8.5 and 8.6, and described below, will show that there are, in general, more individual items which the instructed group scores consistently well on than the exposure group for both grammar and vocabulary. This section will report on general results only; a discussion of individual items and why they may have received high or low scores is contained in a following section.

Since not all items were tested at each testing time and since in many cases equal scores were found for both groups on an individual item, it is important here to operationalize what it means for a group to score better across all three testing times. For one group to show better results on a particular item, a convenient and conservative measure is that it must have the higher score in two of the three testing times even if the item was tested only at two testing times, or if a tie score is reported for the two groups at one or more of the testing times. Thus, for example, if at T1 the instruction group had superior results, then at T2 the item was not tested, and at T3 there were tied results between the instruction and exposure group, this particular item would not be counted in favor of either the instructed group or the exposure group. Also in the case of an item not being tested at one time and the high scores being shared between the instruction and exposure group on the following two times, the results would obviously favor neither group.

Table 8.5. Round One and Round Two target grammar items sorted by which group consistently scored higher on them across the three testing times.

Round 1 Instructed group	Exposure group	Results inconclusive or tied
Cleft sent. w/ subject focus		possessive + gerund
Pseudo-cleft emph. verbs		*get* passive
"*give* a verb"		*have* a verb
Round 2 **Instructed group**	**Exposure group**	**Results inconclusive or tied**
Pseudo-cleft *is...is*		*recommend* + gerund
despite + gerund		adverb + inversion
conditional + inversion		Judgment verb + *be*

The results for grammar for Round One and Round Two are presented in Table 8.5. For grammar, across the two rounds, there were no cases where the exposure group scored higher on any item at two of the three testing times. The instruction group scored better on three

of the six constructions in Round One, and three of the six in Round Two, with the results for the remaining items being inconclusive. Thus for grammar, the instructed group does score consistently better on a number of individual items across time as compared to the exposure group.

Table 8.6 contains the results for vocabulary where similar, but not as conclusive results are found. For Round One, five of the 12 vocabulary items received more consistent high scores by the instructed group, while seven items had inconclusive results. In Round Two, three items received the highest scores by instructed group and three by the exposure group, the rest of the results being inconclusive. Thus, while the majority of results are inconclusive, the instructed group scores better than the exposure group in Round One, and equally well in Round Two.

Table 8.6. Round One and Round Two target vocabulary items sorted by which group consistently scored higher on them across the three testing times.

Round 1		
Instructed group	Exposure group	Results inconclusive or tied
deliberately		*betray*
expose		*condense*
principle		*deteriorate*
depict		*profound*
propensity		*succession*
		sustain
		trigger
Round 2		
Instructed group	Exposure group	Results inconclusive or tied
ambiguous	*bulk*	*compel*
contemplate	*tangible*	*consent*
flurry	*enhance*	*cumbersome*
		evoke
		incentive
		intrinsically

Thus, looking at the results of individual items across time and across grammar and vocabulary, a generalization can be made that while many of the results are inconclusive, the instructed group certainly does outperform the exposure group, particularly in the case of grammar, and also in the case of vocabulary, though to a lesser degree.

8.2. INDIVIDUAL ITEMS WHICH SHOW THE LARGEST DIFFERENCES BETWEEN GROUPS

Another way of comparing results on target items is to compare them in terms of which ones show the greatest difference in scores between the instruction and exposure groups. That is, while the previous analysis simply looked at which group did better on which items, it is also possible to look at those items on which one group clearly outperformed the other. While there are many inconclusive results concerning individual items, as shown in the above tables, looking at cases where the difference between the groups is large, may allow insights into which items posed the greatest challenge for noticing when comparing the exposure group to the instruction group. The question, then, being looked at here is this: on which items did instruction have the greatest impact as shown by scores with the greatest difference between the exposure and instruction groups?

Beginning with the results for grammar, the difference in scores between the instruction group and the exposure group can be found in Table 8.7. Results are presented from smallest to the greatest differences. These data obviously reflect the fluctuating scores across testing times, as has been noted in several places above, thus it will not be surprising to find variable results, particularly in Round One at testing time 2, as has been discussed before. Concerning Round One grammar, the pseudo-cleft and cleft sentences can be considered to be the structures which show the greatest difference between the instructed group and the exposure group, and, thus, show the greatest impact of instruction on noticing. The pseudo-cleft sentence construction shows the greatest difference between the two groups at T1 and T2 and, then, is the second greatest difference at T3. The cleft sentence is ranked second at T1, quite low at T2 (actually in favor of the exposure group at this point), and, at T3, shows the same score as for the pseudo-cleft. The drop in rank at T2 is due to an unusually low score by the instruction group and a somewhat higher score by the exposure group. Thus, somewhat discounting the drop in scores for the cleft sentence at T2, there is reasonable evidence that the structures whose noticing is most affected by instruction are the cleft and pseudo constructions. For the other structures, they either remain at a stable low to mid level, such as the "*give* a verb" construction, with low to mid-range scores consistently in favor of the instruction group, or there are structures that show quite large variation, such as the "*have* a verb" construction, which received the lowest score at T1, and yet the highest at T3, as seen from the point of view of the instructed group.

Concerning grammar in Round Two, a similar picture can be found. Again, as noted before, the expected variation is found, yet taking this into account, there are structures for which the instructed group fairly consistently scores larger than the exposure group. Again, the pseudo-cleft construction, this time the "*is...is*" construction, shows large differences between the two groups at T1 and smaller, though still second largest, at T3. At T2 though, there is again the situation where both groups showed low scores, in this case, the same (14.29) thus resulting in a score of 0 for the difference between the two groups. Thus, again discounting the T2 results, the cleft construction shows one of the largest differences between the two groups. Also showing a large difference is the "*despite* + gerund" construction, which

shows the largest differences at T2 and T3, and smaller differences, though still in favor of the instruction group at T1. A third type of category which qualifies marginally for inclusion in the top group would be those involving inversion, the "conditional + inversion" and "adverb + inversion" constructions. These each only appear twice in the data due to being removed for control purposes, though, taken together, each one makes one appearance as showing the second highest differentiation between the instructed and uninstructed groups. At the same time, they also score on the lower end of the scale as well, with the "conditional + inversion" being the lowest, though still positive, scorer at T1, and the "adverb + inversion" construction at T3 showing a negative score, thus being in favor of the exposure group. Therefore, the results for these constructions are quite mixed, though taken together they do show large differences between the exposure and instruction group on at least two occasions.

Table 8.7. The difference between the instructed group and exposure group scores on individual target grammar items.

Round 1					
1.1	"*have* a verb"	"*give* a verb"	possessive + gerund	Cleft sent. w/ subject focus	Pseudo-cleft emph. verbs
	–7.69	15.38	23.08	38.46	69.23
1.2	possessive + gerund	Cleft sent. w/ subject focus	"*give* a verb"	*get* passive	Pseudo-cleft emph. verbs
	–15.38	–7.69	7.69	7.69	23.08
1.3	get passive	"*give* a verb"	Cleft sent. w/ subject focus	Pseudo-cleft emph. verbs	"*have* a verb"
	7.69	7.69	15.38	15.38	23.08
Round 2					
2.1	conditional + inversion	*despite* + gerund	*recommend* + gerund	adverb + inversion	Pseudo-cleft *is...is*
	14.29	28.57	28.57	42.86	42.86
2.2	Pseudo-cleft *is...is*	*recommend* + gerund	Judgment verb + *be*	conditional + inversion	*despite* + gerund
	0	0	0	28.57	42.86
2.3	Judgment verb + *be*	*recommend* + gerund	adverb + inversion	Pseudo-cleft *is...is*	*despite* + gerund
	–28.57	–14.29	–14.29	14.29"	28.57
Note: a negative score signifies a difference in favor of the exposure group, while a positive score means a difference in favor of the instruction group.					

For grammar, then, an argument can be made that the cleft and pseudo-cleft constructions, the "*despite* + gerund" construction, and, more speculatively and with less evidence, the constructions involving inversion, all show the greatest impact of instruction in making them salient and thus encouraging noticing. But why do these particular structures show the strongest results? There is a different potential answer for each construction. The cleft and

pseudo-cleft constructions are the ones which involve the greatest change in sentence structure of all the constructions, and thus it would be expected that this would make the construction salient for all participants. For instructed participants, this natural saliency combined with the effects of instruction ensured relatively high scores on these items. For uninstructed participants, some rather high results were also achieved on these structures, for example the pseudo-cleft construction in Round One at T2, where the uninstructed group scored 30.77, and the pseudo-cleft construction in Round Two, where at T1 and T2 this was the highest score achieved by the group. Yet, while saliency may make these scores higher, the fact that comprehension of sentences with these constructions is not impaired by not knowing the constructions may have lessened the tendency for exposure group participants to focus their attention on these constructions. That is, if participants did not attend to the structures because they appeared more salient to them, it was unlikely there was any breakdown in comprehension which would require them to attend to these constructions. Thus while these constructions could naturally be equally salient to both groups, the effects of instruction, on the one hand, and lack of need to focus on the construction for reasons of comprehension on the other, served to tip the balance in favor of the instructed group.

The second construction to consider is the "*despite* + gerund" construction. The explanation for the better results for the instructed group is relatively straightforward. The possibility of using this construction was basically unknown by the participants of the study, having learned, probably early on, that *despite* could be followed by a noun. In discussing this structure with the head teacher for all of the students, she explained that this structure was one which they had not been taught yet. For those participants who were in the instruction group, it was new information that this construction was possible, and thus it was made salient for them in the text, though the structure itself is certainly not as unusual as the cleft constructions and thus would appear to be less naturally salient. Therefore, when these participants encountered the item in the text, they may not have noticed it there, and furthermore, when seeing the item again on the noticing test in a forced-choice format, paired up with a very familiar construction, it may have seemed a safe choice for them to choose the more familiar expression, or to choose a "Guess" response rather than a "Remember" response.

Finally, if the inversion constructions do qualify as ones which the instruction group scores better on than the exposure group, the reason most certainly has to do with the fact that inversion is naturally not a salient feature and could easily go unnoticed. Instruction would then have the role of focusing students' attention on this hard to notice structure, which would pay off in contrast to those attempting to notice a small change in sentence structure which has no impact on meaning.

Scores for vocabulary which show the difference between instructed and uninstructed participants are given in Table 8.8. Clear patterns concerning vocabulary are less forthcoming than for grammar. For Round One, three words score in the higher ranges at least on two occasions: *propensity*, *deliberately*, and *expose to*. The item *propensity* shows the third highest difference between the instruction and exposure groups at T1 and T3, and was not tested at T2. The item *deliberately* scores second highest at T1 and T2, yet falls off to show a difference in favor of the exposure group at T2, though again this appears to be due to an unusual decrease

in scores for the instruction group followed by an increase for the exposure group. The third item to show a high difference between instruction and exposure groups is *expose to*, which shows the largest difference at T1, the second largest at T2, and then the difference falls off to be at the top of the lower range, yet still in favor of the instruction group. For Round Two, the results are more mixed with some wide variation shown. For example, the item showing the greatest difference in favor of the instruction group at T1, *bulk*, shows the greatest difference in favor of the exposure group at T2 and T3. In general, the results are weak and inconsistent when compared across the three testing times. The two items which do come closest to showing a consistent difference between exposure and instruction are *contemplate* and *ambiguous*, both of which are tied for showing the third highest difference between instruction and exposure at T1, and the highest differences at T2. It should be pointed out, though, that at T2 all of the differences in favor of the instruction group are quite small.

Table 8.8. The difference between the instructed group and exposure group scores on individual target vocabulary items.

Round 1									
1.1	*sustain*	*principle*	*deteriorate*	*succesion*	*betray*	*profound*	*propensity*	*deliberately*	*expose*
	0	7.69	15.38	15.38	15.38	23.08	30.77	38.46	53.85
1.2	*profound*	*deteriorate*	*deliberately*	*trigger*	*principle*	*condense*	*succession*	*expose*	*depict*
	−38.46	−15.38	−15.38	−7.69	−7.69	0	0	15.38	23.08
1.3	*betray*	*condense*	*sustain*	*depict*	*expose*	*principle*	*propensity*	*deliberately*	*trigger*
	0	0	0	7.69	7.69	15.38	30.77	38.46	46.15
Round 2									
2.1	*tangible*	*evoke*	*intrinsically*	*flurry*	*compel*	*ambiguous*	*contemplate*	*consent*	*bulk*
	−14.29	−14.29	0	14.29	14.29	14.29	14.29	28.57	28.57
2.2	*bulk*	*cumbersome*	*enhance*	*compel*	*evoke*	*incentive*	*tangible*	*ambiguous*	*contemplate*
	−28.57	−28.57	−14.29	0	0	0	14.29	14.29	14.29
2.3	*bulk*	*tangible*	*compel*	*enhance*	*incentive*	*consent*	*cumbersome*	*intrinsically*	*flurry*
	−42.86	−28.57	−14.29	−14.29	−14.29	0	0	14.29	28.57

Note: a negative score signifies a difference in favor of the exposure group, while a positive score means a difference in favor of the instruction group.

Thus, taking the variable results into account, there are five words from the two rounds which show the greatest difference between the exposure and instruction group and hence the greatest effect of instruction: *propensity*, *deliberately*, *expose to*, *ambiguous*, and *contemplate*. These five words represent all classes of words used as target items, so word class can be ruled out

as an explanation as to why these should show the strongest results. Though the explanation for these results is surely a complex one, one element which may play some role is word length. With the exception of *expose to*, each of the remaining four words comes from the group of the eight longest words used as target items. The item *expose to* itself is unusual due to being the only verb with a preposition. While length might seem to make these words easier to notice, length might also have the effect of encouraging participants to skip over the words as they may be somewhat difficult to pronounce in comparison to other words used in the texts. Thus, using a strategy of skipping over difficult words might lead some students to not focus attention on them. For instructed learners, on the other hand, these words would have been familiar, and seeing a newly learned unusual or long word in a new context may promote some participants to direct their attention to the word.

8.3. INDIVIDUAL ITEMS WHICH SHOW THE HIGHEST SCORES OVERALL

Still another way of looking at the individual data is to look for items which receive the highest scores for noticing. Thus, rather than looking at which target items showed the greatest differences between the two groups, as was done in the previous section, this analysis will look at each group separately to see which items received the highest noticing scores.

Tables 8.9 through 8.12 show the rank-ordered grammar and vocabulary items for each group at each testing time, ranked from the highest noticing score to the lowest. These tables will be discussed both in this section and the following section, and thus not all of the information present on them, for example the lines connecting up target items, is relevant for this present discussion.

Tables 8.9 and 8.10 present the data for grammar for Rounds One and Two. In Round One, the "*give* a verb" construction received the highest noticing score across all three testing times, showing comparatively high scores in both the instructed and exposure condition. In Round Two, at T1 the pseudo-cleft construction is the high scorer for the instructed group and also for the exposure group, though the "*despite* + gerund" construction received the same score in the latter case. At T2, again in Round Two, the conditional + inversion received the highest score in the instructed group, and in the exposure group there was a three-way tie between that construction, the pseudo-cleft, and judgment verb + *be* construction. It needs to be noted, though, that the high scores in the exposure group are actually rather low in comparison to the instruction group. At T3, the pseudo-cleft construction again received the highest score from both groups.

Thus, for grammar, the target items which receive the highest scores were the pseudo-cleft, conditional + inversion, "*despite* + gerund", "*give* a verb", and judgment verb + *be* constructions. The first three constructions were discussed above, and the same arguments apply here as to why these items received the noticing scores they did, especially when noting that the "high" scores for the exposure group are actually very low scores when compared to the

Table 8.9. Target grammar items ranked from highest to lowest noticing score for Round One instruction and exposure groups

T1		T2		T3	
Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
"*give* a verb"	"*give* a verb"	"*give* a verb"	"*give* a verb"	"*give* a verb"	"*give* a verb"
92.31	76.92	69.23	61.54	61.54	53.85
Pseudo-cleft emph. verbs	"*have* a verb"	Pseudo-cleft emph. verbs	Pseudo-cleft emph. verbs	Cleft sent. w/ subject focus	Cleft sent. w/ subject focus
46.15	7.69	53.85	30.77	23.08	7.69
Cleft sent. w/ subject focus	possessive + gerund	Cleft sent. w/ subject focus	Cleft sent. w/ subject focus	"*have* a verb"	*get* passive
30.77	0.00	7.69	15.38	15.38	7.69
possessive + gerund	Cleft sent. w/ subject focus	possessive + gerund	possessive + gerund	*get* passive	"*have* a verb"
23.08	–7.69	–7.69	7.69	15.38	–7.69
"*have* a verb"	Pseudo-cleft emph. verbs	"*get* passive"	"*get* passive"	Pseudo-cleft emph. verbs	Pseudo-cleft emph. verbs
0.00	–23.08	–7.69	-15.38	7.69	–7.69

instruction group, except for the case of the pseudo-cleft at T3. The judgment verb + *be* construction seems unusual to be considered in the high scoring category, as it does not receive high scores, save for T3, where it received a higher score. Therefore, it is possible that its inclusion in the high scoring group is more a result of the three-way tie which occurred in the exposure group at T2 than any particular characteristic of the construction. On the other hand, the "*give* a verb" construction received high scores for both groups across each testing time. The very high scores for both the instruction and exposure group could be due to the fact that the construction involved the use of a known lexical item in an unusual way. That is, although participants were almost surely familiar with *give* in its usual two-object construction, using *give* + indirect object + deverbal noun object could have seemed unusual and therefore noteworthy for participants. Thus, for both groups, high scores are achieved across each testing time, with the consistently larger scores by the instructed group being evidence of an effect of instruction.

Table 8.10. Target grammar items ranked from highest to lowest noticing score for Round Two instruction and exposure groups

T1		T2		T3	
Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
Pseudo-cleft *is...is* 71.43	Pseudo-cleft *is...is* 28.57	conditional + inversion 42.86	conditional + inversion 14.29	Pseudo-cleft *is...is* 57.14	Pseudo-cleft *is...is* 42.86
despite + gerund 57.14	*despite* + gerund 28.57	*despite* + gerund 28.57	Pseudo-cleft *is...is* 14.29	*despite* + gerund 28.57	Judgment verb + *be* 28.57
adverb + inversion 42.86	conditional + inversion 14.29	Pseudo-cleft *is...is* 14.29	Judgment verb + *be* 14.29	*recommend* + gerund 0.00	*recommend* + gerund 14.29
conditional + inversion 28.57	adverb + inversion 0.00	Judgment verb + *be* 14.29	*recommend* + gerund 0.00	adverb + inversion 0.00	adverb + inversion 14.29
recommend + gerund 14.29	*recommend* + gerund –14.29	*recommend* + gerund 0.00	*despite* + gerund -14.29	Judgment verb + *be* 0.00	*despite* + gerund 0.00

Vocabulary for Rounds One and Two is presented in Tables 8.11 and 8.12. In Round One, the items which received the highest scores for noticing were *succession*, *trigger*, and *principle*, which were the highest for both groups across the three testing times, and *profound* and *deteriorate*, which scored the same as *succession* for the instructed group at T1 and the same as *trigger* for the exposure group at T2. Each of the five items noted received rather high scores from each group. In Round Two, the items which received the highest score were *flurry*, *tangible*, *bulk*, and *cumbersome,* with this last item having the same score as *bulk* for the exposure group at T2, and, again, here each item received rather high scores.

Table 8.11. Target vocabulary items ranked from lowest to highest noticing score for the instruction and exposure group.

T1		T2		T3	
Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
succession	*succession*	*trigger*	*trigger*	*principle*	*principle*
69.23	53.85	69.23	76.92	76.92	61.54
profound	*profound*	*principle*	*deteriorate*	*deliberately*	*deliberately*
69.23	46.15	61.54	76.92	69.23	30.77
deliberately	*sustain*	*deteriorate*	*principle*	*trigger*	*betray*
61.54	46.15	61.54	69.23	61.54	30.77
expose	*principle*	*expose*	*profound*	*propensity*	*expose*
53.85	46.15	38.46	61.54	53.85	30.77
principle	*deteriorate*	*deliberately*	*deliberately*	*expose*	*depict*
53.85	38.46	30.77	46.15	38.46	23.08
deteriorate	*betray*	*depict*	*expose*	*betray*	*propensity*
53.85	30.77	30.77	23.08	30.77	23.08
betray	*deliberately*	*profound*	*condense*	*depict*	*trigger*
46.15	23.08	23.08	15.38	30.77	15.38
sustain	*propensity*	*condense*	*depict*	*condense*	*condense*
46.15	7.69	15.38	7.69	15.38	15.38
propensity	*expose*	*succession*	*succession*	*sustain*	*sustain*
38.46	0	7.69	7.69	15.38	15.38

One explanation as to why these words discussed above receive the highest noticing scores could be their unusual spelling, in which there are examples of double letters and unusual combinations of letters in words. This could have made the words more salient due to these features, or it could have made some of the words difficult to pronounce and hence more likely to be more memorable if time was spent on trying to sound them out.

Table 8.12. Target vocabulary items ranked from lowest to highest for Round Two for the instruction and exposure groups.

T1		T2		T3	
Inst.	Exp.	Inst.	Exp.	Inst.	Exp.
flurry	*flurry*	*tangible*	*bulk*	*tangible*	*tangible*
85.71	71.43	57.14	57.14	57.14	85.71
bulk	*tangible*	*ambiguous*	*cumbersome*	*flurry*	*bulk*
71.43	57.14	42.86	57.14	57.14	85.71
ambiguous	*bulk*	*evoke*	*evoke*	*bulk*	*cumbersome*
57.14	42.86	42.86	42.86	42.86	42.86
consent	*ambiguous*	*bulk*	*tangible*	*cumbersome*	*enhance*
42.86	42.86	28.57	42.86	42.86	28.57
tangible	*evoke*	*enhance*	*enhance*	*intrinsically*	*flurry*
42.86	42.86	28.57	42.86	42.86	28.57
evoke	*consent*	*compel*	*compel*	*enhance*	*intrinsically*
28.57	14.29	28.57	28.57	14.29	28.57
compel	*compel*	*cumbersome*	*ambiguous*	*compel*	*compel*
28.57	14.29	28.57	28.57	0.00	14.29
contemplate	*contemplate*	*incentive*	*incentive*	*consent*	*consent*
28.57	14.29	14.29	14.29	0.00	0.00
intrinsically	*intrinsically*	*contemplate*	*contemplate*	*incentive*	*incentive*
14.29	14.29	14.29	0.00	-14.29	0.00

8.4. COMPARISON OF ACCURACY ORDERS FOR INDIVIDUAL ITEMS BETWEEN GROUPS

A further different way of looking at the same data is to look at the order of items for each group at each time according to how accurately they are noticed. In the previous section, the highest scoring items were discussed, and here I want to focus on relative ordering of all items. The point which will be made here is that the ordering of items by accuracy is roughly similar between the instruction group and exposure group at each testing time, though with some notable exceptions.

Data for this section were presented in Tables 8.9 through 8.12 above. The similar ordering of items between the instruction and exposure group is shown by lines connecting the items. Two important issues need to be mentioned here in explaining how the ordering of the items was determined. First, items in each column were considered to be in the same order if they were next to each other or immediately above or below each other. In the case where two possible orders could be chosen, as can be seen in Round One at T1 between the cleft sentence and possessive + gerund, only one order was chosen and marked with a solid line,

while the other possible order was marked with a dashed line. In this case though, only one of the two pairs was tallied as being in the same order. Second, and more importantly, when within one group if items had the same score, they were freely moved around so as to make the best order of accuracy possible. Thus, it is important to note at the outset that these results are as dependent on an actual accuracy order as they are on the possibility of multiple repeated scores within groups giving the flexibility which allows for the adjusting and matching an order of accuracy across groups. Thus, to the degree that multiple items receive the same score within groups, it will be easier to find a similar accuracy order between groups. This poses strong limitations on the validity of this analysis, but it is possible to cautiously speculate from the results that are found.

What is immediately apparent when looking at the accuracy orders is that they are quite similar between the instruction and exposure groups at each time. For grammar at Round One at T1, T2, and T3, there are two, five, and four similarities in order, respectively, from a possible total of five at each testing time. That is, for example at T2, each of the items for both groups are in the same order, thus there are five similarities, while at T1 two items can be ordered across the two groups. For Round Two grammar, the results are four, four, and three similarities in order across the testing times. For vocabulary in Round One, the results are similar, with six, five, and six similarities out of a possible nine across the three testing times, respectively. For Round Two there are seven, five, and seven similarities out of a possible nine across the three testing times, respectively.

What this suggests is that even though the groups received different treatments and that there are differences between the results of the groups — where in many cases the instruction group outperforms the exposure group — still, all of these results are, in many cases, being influenced by some common underlying factors. The source of these factors must be a complex mixture of the way items are positioned in the text, their role in the narrative, their saliency due to their structure, and participants' own experience of these items based on their proficiency and previous knowledge. That is, while in many cases instruction may give an advantage to participants — which can be seen as an increase in score between the instructed and exposure group — that boost given by instruction is added on to whatever factors conspire to make an item noticeable. Looking at a graph of the results from Round One, T1 for vocabulary can make this point clearer. This is presented in Figure 8.1, where items have been ordered to show their possible common order. In Figure 8.1, the first six items can be placed in a common order showing that the increase in score gained by the exposure group is added on to a possible constant which is present in the saliency of the items themselves, as discussed above. The last three items do not fit into the order.

It is now important to look at the one case across the two rounds which is an exception, that is Round One grammar at T1, shown in Table 8.9. Here, only two similarities can be found between the orderings of items where at other testing times more similarities are found. This exception can be explained by the fact that this testing time shows that greatest mean differences between two groups (27.69 points) found in any of the data for grammar or vocabulary across both rounds. One speculative answer as to why these data are an exception

could be that the effects of instruction are strong enough to wash out most of the effects of the natural saliency of the target items.

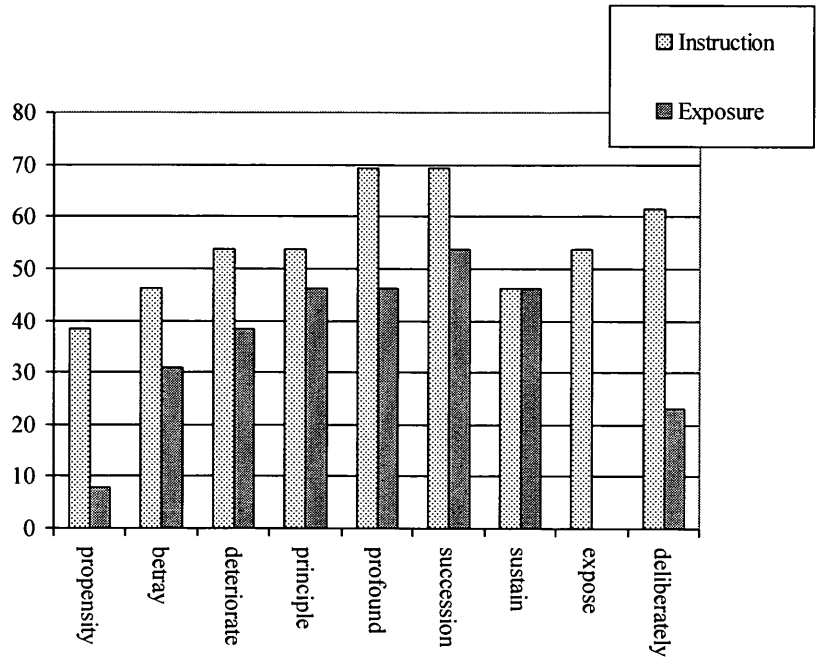


Figure 8.1. Round One, T1 percentages for target item vocabulary noticing for the instruction and exposure groups.

Again, it is important to underscore the fact that this line of analysis is largely speculative and based on what could turn out to be a rather weak method of creating accuracy orders between groups.

8.5. SUMMARY

This chapter has looked at patterns of scores on individual items. Several points have been made. First, it was pointed out that the overall results show the same pattern reported in previous sections of a strong showing for effects of instruction at the first testing time, followed by rather mixed results at later testing times. It was also pointed out that if individual items are the focus of analysis over time, then the instructed group has better and more consistent results compared to the exposure group in most cases.

Second, the differences in scores between the instructed and exposure groups were looked at to see which items benefited the most from the effects of instruction. Results showed a number of vocabulary and grammar items for which instruction clearly promoted greater

noticing. Possible explanations for this could be argued for grammar, though weaker arguments were made for vocabulary as well.

Third, individual items were analyzed within groups to see which ones were noticed to a greater degree than others. Again, certain grammar and vocabulary items could be identified, and stronger explanations could be delivered in the case of grammar than vocabulary.

Finally, accuracy orders for noticing grammar and vocabulary were speculatively established which showed that, across most testing times, the orders of accuracy between the instructed and uninstructed groups have more in common than not. The explanation proposed for this concerned the underlying positive effect that the natural saliency of items has on noticing.

CHAPTER 9. GENERAL DISCUSSION

The previous chapters presented and discussed data concerning noticing and learning from the perspective of the comparison of groups, individuals, and individual structures. In this section, I will look at these results in general, drawing on each of those sections to get a composite picture of what can be said concerning the noticing of grammar and vocabulary in input and the relationship to learning, as demonstrated by these data. Secondly, in various places in the section I will address the issue of consciousness and the use of explicit knowledge in second language learning and discuss implications that the data presented in this study can have. Furthermore, I will critique the test for noticing and address the question of the viability of this kind of instrument for use in testing noticing.

Again, it is necessary to mention at the outset that the majority of the comparisons made in the previous chapters were found to be statistically non-significant and contain a high degree of variability. This is due in part to the very small sample sizes used, with, in the more extreme case, groups of seven participants being involved in comparisons in Round Two. Added to this is the perhaps difficult decision that the noticing test asked participants to make concerning their states of consciousness of memories, which, despite pre-test training and efforts to control response bias to some degree, may have introduced a larger degree of variability in responses. This issue will be taken up below when the noticing test itself is discussed in more detail. Added to this is the inherent variability introduced when doing classroom-based research. There are interruptions and distractions which delay classes, necessitate the moving of testing times, and, generally, compete for the attention of the student participants. Therefore, while the non-significant results cannot be overlooked, they may in part be attributed to the factors mentioned above. Finally, what was focused on in the previous chapters will be emphasized here again: despite overall non-significant results, the general patterns that emerge are many times in support of the hypotheses and general position that this project is investigating.

9.1. NOTICING GRAMMAR AND VOCABULARY IN INPUT

As was noted in the literature review in Chapter 2, noticing has recently been ascribed a variety of important roles in second language learning, from being the necessary element in the learning process (e.g. Schmidt, 1994) to being one of several important factors in learning, which, in this case, plays a role in moving learners to a higher stage of language processing through the noticing of gaps between one's output and the input (Swain, 1995; Swain &

Lapkin, 1995). One of the main points of this study has been to establish that there is a relationship between instruction and noticing forms in input. Therefore, it is important to draw some general conclusions from the previous four chapters and note the degree to which a relationship between instruction and noticing has been established through these data.

In the following sections three areas will be discussed: the impact of instruction on noticing, the difference between the noticing of grammar and vocabulary, and the evaluation of participants' responses to target items encountered in the noticing texts.

9.1.1. THE LIMITS OF THE IMPACT OF INSTRUCTION ON NOTICING

The preceding chapters show that instruction can generally be seen to have a positive impact on noticing for both grammar and vocabulary, particularly at the first testing time, T1, where in each case in both Round One and Round Two the instructed group had the highest scores for noticing of any group in all testing times. These results were similarly expressed in scores for individual items where at T1, as was shown in Tables 8.1 through 8.4, across both grammar and vocabulary, the instructed group showed higher scores on every individual item for Round One vocabulary and Round Two grammar, and the highest scores on all but one item for Round One grammar and highest or equal scores for all but one item in Round Two vocabulary. These data show quite consistent results that instruction does have a relatively strong impact on noticing on the immediate post-test. And it is important to point out here that the "immediate" post-test was between four to five days after instruction, depending on the group, from the end of instruction. So the results are durable, at least to five days after instruction.

Noticing induced by the specific type of instruction used in this study thus had rather good short-term durability, but did decrease sometime between the 5th and the 11th day of instruction, when testing time two began. After this point, the results become more variable and all scores are consistently lower than those for the instruction group at T1. Thus, it appears that, below a certain level, both the instruction and exposure group have more of an equal chance in noticing target items, the result of which is consistently less distance between the means for both groups, and some instances where the exposure group outperforms the instruction group. This pattern was once again seen in the results for the individual items, shown in Tables 8.1 to 8.4, where, after the instruction group scores better on nearly every grammar or vocabulary item at T1, the exposure group scores better or equally well on many items, particularly in vocabulary, at later testing times.

Therefore, the case can be made that although there is a large amount of variability in the data when looked at as a whole, instruction does have an impact on noticing, and the variable results that occur at testing times two and three really bring into question the issue of the durability of an effect of instruction on noticing, rather than the existence of such an effect, which can be established at T1.

Yet, why do these variable results exist after T1? Clearly, the instruction group is showing a loss of retention of memories of the instructed items, a loss which follows the same pattern found for memory in general, where there is steep initial drop-off in ability to remember fol-

lowed by slower decay over time (Anderson, 1995:234). Examples of this same power law of forgetting are shown in other studies which use the Remember/Know paradigm, where the durability of Remember responses in word recognition tasks has been demonstrated to decline sharply within a period of one week after exposure (Gardiner & Java, 1991). Yet, while this explains the decay in results for the instructed group, it may not explain the ability of the exposure group to do as well as it does.

There are several possible explanations for why the exposure group does as well as it does. First, the results of the exposure group could just be showing the effects of exposure without noticing. This explanation seems unlikely, though. The exposure that these participants received was minimal, amounting to reading two passages containing all of the target items while their attention was being divided between meaning and form through task directions at the time of reading and enhanced input of the target items for which they served as the instruction group. That is, their attention was generally being directed to other areas than the target exposure group items. This small amount of exposure can be contrasted with studies using "input flood", for example Trahey & White (1993), where French learners of English were exposed to hundreds of examples of the target structure, adverb placement, over a short period of time, showing the benefit of participants learning what possibilities exist in English for the position of adverbs. Although the focus of the Trahey & White study was learning and not noticing, it can serve as a contrasting example showing the amount of exposure which may be needed to achieve a positive effect on learners. Thus, concerning the present study, it seems unlikely that this small amount of exposure that the exposure group received would have influenced participants' scores to any great degree, although this cannot be ruled out.

Another possibility is that participants from one group communicated with participants from the opposite group concerning what was being focused on in the instructional sessions. Although it is impossible to rule this out — and, in fact, participants were overheard in the course of the study discussing what it was the opposite group was learning — the effects of this would most likely have influenced only isolated individuals and not an entire group.

What seems a likely explanation for the rather similar results at T2 and T3 for the instruction and exposure group is that this level of noticing roughly represents the natural amount of material which can be noticed and retained within the duration of the noticing test, based on the natural saliency of the target items as they appeared in input. Effects for instruction can still be seen, but the differences between the instruction and exposure group are smaller than at T1. Thus, noticing is influenced by collective influences of instruction and also the natural saliency of items, with the effects of instruction wearing off over time, and the natural saliency of items remaining constant. This interpretation of complementary influences which affect noticing is supported in a review and discussion of the model of noticing presented by Skehan (1998:48-52), based mainly on Schmidt (1990). Four factors are discussed as having an influence on noticing. Skehan lists two characteristics of input which have an impact on noticing: input qualities, including frequency and saliency of items, and focused input through instruction or tasks. Furthermore, task demands and factors specific to individuals, such as processing capacity and readiness to receive input, also influence noticing. For the noticing test, task demands were held constant for each group and thus may not be relevant

here (although see the discussion below). Individual differences have been shown to exist and were discussed in Chapter 7, but it seems unlikely that individual differences could account for the variable results found at T2 and T3. The two factors left, then, are the input factors of saliency and the focusing of input through instruction — the two factors identified as most likely causing the results found in these data.

This position is supported by evidence presented in the discussion of the results of individual target items in Chapter 8. Here, a comparison of accuracy orders for target items between the instruction group and exposure group for grammar and vocabulary were presented in Tables 8.11 and 8.12. These tables showed that while at T1 there were rather different accuracy orders between the instruction and exposure groups for target items, at T2 and T3 the accuracy orders were generally more similar. The claim was made that the different accuracy orders for noticing at T1 were caused by the effects of instruction, overriding any saliency effects found in the individual items, while at T2 and T3 the more similar accuracy orders suggest that it is something in the input itself that is causing the scores to be roughly similar — or at least more similar than at T1, when stronger effects of instruction were found.

Another factor which is likely to have had a strong influence on the scores achieved by the exposure group was that the directions the noticing test required participants to focus on form. As was mentioned above, the task demands were the same for both groups, and what this means is that while the instruction group was naturally focused on form, the exposure group may have been unnaturally influenced to focus on form due to the demands of the noticing test. This is a potential weakness of the testing format itself, which will be discussed below in Section 9.3. The result is that an additional element of an orientation towards focusing on form must be added to the potential saliency of target items as an explanation as to why the exposure group did as well as it did.

Speculative evidence for the effects of the orientation of participants towards form can be provided by the data presented for Round One noticing in Section 5.7.1.1. These results show that in Round One, between T1 and T2 there is an unusual reversal in the tendencies of results where exposure scores rose for both groups and instruction scores dipped for both groups. Since the instruction groups served simultaneously as exposure groups, it was claimed that this could have happened as a result of groups of participants adopting a strategy to focus on the area which they had not received instruction in, but were tested on in the previous test. Thus, if this analysis is correct, this example can provide evidence for the effects of participants' orientation on what was noticed.

9.1.2. THE DIFFERENTIAL EFFECT OF INSTRUCTION ON GRAMMAR AND VOCABULARY NOTICING

Even within variability and lower level results just discussed, patterns can be seen which point to the positive effects of instruction. Here and in general, instruction can be seen to benefit grammar to a greater degree than vocabulary. These results, already seen in Chapter 5 in Tables 5.3 and 5.4, support the differential effects of instruction on grammar and vocabulary, where instruction in grammar produced more consistent positive results when compared to

instruction in vocabulary. In five out of six cases across the two data collection rounds the instruction group scores better than the exposure group for grammar, but for vocabulary the instruction group scores better on only three out of six cases. That is, even though at T2 and T3 the mean scores are lower, closer together and not statistically significant in either Rounds One or Two, a general pattern showing an advantage for grammar instruction can be found. Put another way, there appears to be a greater need for grammar instruction to promote the noticing of grammar than there is for vocabulary instruction to promote vocabulary noticing if results across all testing times are considered.

Evidence for the superior effects of grammar instruction is also found through the analysis of individual items presented in Chapter 8. Tables 8.5 and 8.6 present data showing the consistency of high scores received on grammar and vocabulary items. Results show that no grammar item consistently received high scores by the exposure group in Rounds One and Two, while three vocabulary items, in Round Two did consistently receive higher scores by the exposure group.

This evidence presented in the previous two paragraphs points to the conclusion that grammar instruction has a more consistent and long lasting effect than vocabulary instruction when compared to an exposure group. This, in turn, suggests that grammar instruction is more necessary for noticing grammar than vocabulary instruction is for noticing vocabulary.

Another way of looking at this same question is to compare noticing scores in the exposure condition for grammar and the exposure condition for vocabulary, as was done in section 5.6. This section presents data comparing the scores for the noticing of grammar with scores for the noticing of vocabulary which was the object of exposure. In each case the scores for the noticing of vocabulary which was the object of exposure are higher than that of grammar, pointing to the fact that vocabulary is simply easier to notice than grammar.

The most likely reason for this is that vocabulary is more salient and surface-level, while grammar, even though while here in this study it relates in some cases to specific and consistent lexical items, is generally concealed behind different lexical items in different situations and can be highly abstract in comparison to vocabulary, which, even if changed due to inflection, remains largely recognizable. Indeed saliency is noted as a key factor concerning what will influence the noticing of items in input (Skehan, 1998; N. Ellis 1994; Gass 1988). Thus grammar instruction is useful in making forms more salient which would normally be more difficult to notice. Vocabulary instruction increases the saliency of lexical items as well, but after the initial effects of instruction have weakened over time, it becomes possible for the exposure group to do as well or better than the instructed group, due to the natural saliency of vocabulary items.

This issue of the saliency of items as being a primary factor in promoting or lessening the noticeability of items is also demonstrated in section 7.2 in the analysis of individual items which showed the greatest difference in noticing between the instructed and exposure group. Here it was suggested that those items which showed the greatest difference between the exposure and instruction groups were those which could be considered less salient.

Gass (1998) also suggests a rationale that would support this view of differential noticing of vocabulary and grammar in her discussion of "apperceived input", that is, input which is re-

lated to previous knowledge which has been stored in memory (p. 206). She suggests that a factor which influences whether or not input becomes integrated is whether or not the input has been processed simply in terms of meaning, or whether it has been processed in terms of syntax (pp. 205–206). This would support the view that grammar noticing may be more likely to have consistent results as the majority of forms used in this study would involve a level of syntactic processing, whereas vocabulary items could be processed as comprehended input or surface phonological forms. Thus where instruction has induced a deeper level of processing, differences between the instruction and exposure groups may continue longer over time.

Thus, several different lines of evidence have pointed to the fact that vocabulary items are naturally easier to notice than grammar items, and that explicit instruction, then, is beneficial in increasing noticing scores of grammar items in the longer term when compared to exposure. Expressed in a different way, this points to the position that *items* (that is, for example, individual words) can be more easily noticed than *rules*, and that explicit instruction is more beneficial in establishing noticing in the case of rules than items. Put in these terms, this gives support to the hypothesis suggested by Hulstijn & De Graaff (1994:105) that instruction will have a greater impact on rules rather than items. Their hypothesis is related to the learning of grammar and vocabulary, but could certainly be extended to the noticing of features as well, for which this current project provides evidence. Central to the support of this hypothesis is the evidence that target vocabulary and grammar which were more salient were likely to be noticed with or without the aid of instruction, while less salient target grammar and vocabulary were more likely to be noticed to a greater degree with the aid of instruction.

9.1.3. EVALUATION OF RESPONSES TO THE NOTICING TEST

In this study, participants were led to focus on form, depending on group, through repeated exposure to forms in reading texts, through discussion of and practice in manipulating these forms through instruction, and through task instructions at the time of testing. The Remember responses produced by this orientation have been focused on since these responses were operationalized as the definition of noticing. In this section the Remember responses will be discussed and evaluated in comparison with Know and Guess responses, which have, up until now, received very little attention in this study. Finally, some tentative general conclusions will be drawn when considering all of the response types.

To review briefly, Remember responses were chosen as the main independent variable for this study because they are defined and used in previous studies as reflecting the episodic memory system, that is, the system which records the events and experiences that people have and involves the conscious recollection of the experience of the context of encoding (Tulving, 1985). For this reason, Remember responses were seen as a possible vehicle for recording participants' conscious experience of and possible reflection on grammar and vocabulary when encountered in input. Furthermore, if a definition of noticing is accepted as conscious registration with rehearsal in working memory (Robinson, 1995b), then Remember responses would be an appropriate way of capturing noticing.

The only way of evaluating whether Remember responses do involve the memory of an event is to use retrospections, as was done in at T1 in Rounds One and Two. These data were presented in Chapter 5, section 5.7.1.2. These data show that participants from both experimental groups were quite able to reflect on their experiences of the target items in input and give, in some cases, quite elaborate details about their thoughts at the time of testing. The purpose of doing the retrospections was to verify that participants were indeed reporting on memories and also to reinforce the instructions for the noticing test. There were a small number of cases where, due to the retrospections, participants changed their answers from a Remember to a Know response or vice versa. Nevertheless, the point is that participants were readily able to report on their experience of the target items at least in some form or another, and thus the Remember responses would seem to be valid measures of episodic memory.

Know responses, on the other hand, according to Tulving (1985), represent semantic memory, where encoding has not occurred of the context of experiencing a particular item or piece of information, thus leaving a person with a strong feeling of knowing that something occurred but without being able to re-experience the actual context. This perspective, too, was borne out through the retrospections, where participants providing Know responses were unable to identify particular aspects of the experience of encoding, yet felt strongly that a particular item had occurred in input.

Guess responses represented what seemed to participants to be random responses, with no feelings of familiarity or knowledge of the items or of the encoding experience present. This also was verified through the retrospections.

Tables 9.1 and 9.2 present the data for the accuracy of each possible response for grammar and vocabulary in both Rounds One and Two. Scores are summarized across all three testing times within each round.[3] In that these are accuracy scores, created by subtracting errors from correct responses, it is possible that negative scores can be found, as has been seen in the data previously. In fact, scores can range from −100, meaning100% errors, to +100, meaning 100% correct answers. These data are, then, a summary of all the responses to the target items presented to participants in the noticing tests across the three testing sessions. The instruction and exposure groups are represented separately in the data. Several patterns can be found in these tables which warrant discussion.

First, it can be seen that Remember responses are always the largest number of all three responses for both grammar and vocabulary in each data collection round. Thus, it can be claimed that the effects of instruction, task directions to focus on form, and the natural salience of target items are more likely to lead to noticing than they are to a strong feeling of familiarity or a correct guess that an item was encountered before. It should be noted here that a pattern of responses like this is typical for studies involving the Remember/Know paradigm What can be seen in these data is that when a correct choice is made, a Remember response, showing noticing, is most likely to be made.

---

[3] The reasons for summarizing these results across the three testing times is a matter of convenience; the same general patterns could be shown at each time.

Table 9.1. A summary of scores for grammar noticing for each of the possible responses across all three testing times.

		Remember	Know	Guess	Total
Round 1					
Instruction					
	M	28.72	2.56	−1.03	30.26
	SD	22.58	18.96	31.12	72.66
Exposure					
	M	13.85	−1.54	−14.87	−2.56
	SD	23.63	27.73	29.55	80.91
Round 2					
Instruction					
	M	26.67	13.33	4.76	44.76
	SD	20.41	19.49	28.92	68.81
Exposure					
	M	12.38	5.71	0	18.1
	SD	23.35	15.64	48.71	87.71

Table 9.2. A summary of scores for vocabulary noticing for each of the possible responses across all three testing times.

		Remember	Know	Guess	Total
Round 1					
Instruction					
	M	45.3	17.66	14.25	77.21
	SD	25.42	18.07	17.11	60.61
Exposure					
	M	34.19	12.82	24.5	71.51
	SD	20.59	11.81	17.03	49.43
Round 2					
Instruction					
	M	34.39	16.93	21.16	72.49
	SD	27.14	22.85	24.04	74.03
Exposure					
	M	34.92	16.93	19.58	71.43
	SD	23.07	18.12	19.91	61.1

These results could be interpreted to mean that different responses simply represent differing levels of confidence in one's answer and not different states of consciousness. This issue was addressed by Gardiner & Java (1990), where recognition memory for words and non-words was tested both by eliciting Remember and Know responses from one group, while two levels of confidence ratings were elicited from the second group. Results for non-words

received higher scores for Know responses than Remember responses. As recognition of non-words was said to rely more on perceptual saliency than on deeper processing, higher Know responses were expected. The results for the confidence ratings did not show a similar interaction, and therefore Gardiner & Java conclude that Remember/Know judgements and confidence ratings are coming from two different sources, and that Remembering and Knowing do represent different states of consciousness. This, then, supports the data from the retrospections collected for this project which provide examples of the different states of consciousness proposed. Ultimately, confidence and explicit memories do correlate, as Gardiner & Java point out (p. 27), but this does not necessarily indicate that there is one underlying system.

The issue of Remember and Know responses was also discussed in relation to noticing and the effect of noticing on learning in sections 5.6.3 and 6.5, respectively. Here, a wider definition of noticing was considered which included both Remember and Know responses. The results of these analyses show that the trends present in the data for Remember responses only were strengthened to some degree when Know responses were added in. Including Know responses puts the analysis on a different theoretical footing, as it has been argued in this study that Remember responses alone represent conscious noticing, while Know responses simply represent a feeling of familiarity without any associated consciousness of the particular items at the point of encoding. Since there are good theoretical reasons for focusing on the Remember responses only, the difference between the results of the two analyses could be explained in terms of a larger amount of data providing more consistent and reliable results, something which might have been achieved with Remember responses alone with more data collected. Nevertheless, the question of whether noticing should include both Remember and Know responses is an important one and is far from answered by the data presented in this study.

Going back to Tables 9.1 and 9.2, three other points can be made when considering all of the possible responses together. First, once again differences between scores for grammar and scores for vocabulary can be detected. In every case across both rounds vocabulary scores are, when considering the total for each response, higher than grammar scores, with grammar scores showing more negative scores, that is, more inaccurate responses than accurate, in three cases. Second, the effects of instruction can again be seen where the total, scores for instructed grammar are always much larger — usually double — than the grammar scores for the exposure group. Finally, for vocabulary, the instructed group still scores higher, but the differences between the groups are trivial. These last two points are reflected again in the sum of the Remember, Know and Guess responses, shown in the right-hand column of Tables 9.1 and 9.2. Here, grammar noticing scores are much higher for the instructed group than the exposure group in each round, while scores for vocabulary noticing were nearly equal across both rounds.

In conclusion, then, two important general points can be made here beyond this discussion of Remember and Know responses representing different states of consciousness. First, in form-focused instruction, task directions to focus on form and the natural saliency of grammar and vocabulary items lead to a surprisingly large number of instances of participants correctly reporting a conscious experience of grammar or vocabulary in input when compared with other possible responses. And second, when all responses are considered together, participants show quite good recognition memory, shown either by Remember or Know responses, for vocabulary in particular and also for grammar that has appeared in input.

9.2. THE RELATIONSHIP BETWEEN NOTICING AND LEARNING

Though the main focus of this project has been the establishment of a relationship between instruction and noticing, the key question that lies behind the issue of noticing is the relationship between consciousness and learning. Results relating to this were presented in Chapter 6, which addressed this issue directly, and Chapter 7, which looked at individual differences.

A conclusion that can be drawn from Chapter 6 is that while general support for a positive relationship between noticing and learning can be found from correlational data, particularly in the case of vocabulary, there are many individual examples where a high degree of learning was found without a corresponding high degree of noticing, thus potentially calling into question a strict interpretation of the noticing hypothesis where items must be noticed in input before they can be learned. These results are somewhat mitigated, as was explained in Chapter 6, by pointing out the relatively small numbers of participants who score in the extreme High learning, Low noticing category when the data for the three testing times are collapsed together in each of the data collection rounds. Further mitigating evidence came from Chapter 7 on individual differences, where, in Tables 7.3 and 7.4, it was shown that the participants who consistently had high scores for noticing in both the instructed and exposure conditions tended to have high learning scores as well. Nevertheless, the results which show exceptions — and there are a number of them depending on how they are looked at — and the somewhat low correlation shown in general between noticing and learning are problematic for the noticing hypothesis, at least as it is expressed in its strongest terms by Schmidt (1990), where noticing is necessary for converting input into intake.

There are several points, though, that need to be considered before drawing too strong a conclusion from the data presented in this study. First, it is necessary to look at the extent of the kind of claim that can be made in this project concerning the relationship between noticing and learning. The claim that can be made here in its strongest form would be that noticing and learning co-occur. No causal relationship can be established here, and so the question of why there has been high learning without high noticing needs to be thought of in terms of why this co-occurrence of the two did not happen. With the small numbers of participants and small numbers of items tested, low reliability is a likely problem that needs to be taken into consideration.

Second, as was pointed out above, these results need to be looked at in terms of whether the learning without noticing occurred in the instructed condition or in the exposure condition. Once this is looked at, results show that only a fraction of the instances of learning without noticing occurred in the exposure condition, the rest occurring in the instruction condition where it is easier to explain these results in terms of instruction training participants to manipulate the forms, and thus show learning while not noticing the forms. Although these cases could pose problems for the hypothesis that instruction promotes noticing, they can also serve to weaken the possibility that these data show that *real* learning, as opposed to the manipulation of form, has happened without the benefit of noticing.

A third point is that the tests of learning that were used in this research were testing the controlled manipulation of form in a non-communicative, exercise-like format, and these tests may not accurately enough model the kind of learning that accompanies noticing. Indeed, one of the points of Schmidt's diary study of learning Portuguese (Schmidt & Frota, 1986) was that it was not until he noticed forms in input that they then began to appear in his spontaneous speech. That is, the payoff in using the explicit knowledge gained through noticing may be that learners begin to use that knowledge in a controlled way, possibly, depending on the learner, involving all of the variation and inaccuracies associated many times with the early stages of learning. The tests of learning used in this research may simply not have been fine-grained enough to measure this kind of learning, particularly in its early stages.

Finally, in further arguing against a strict interpretation of the results of the relationship between noticing and learning shown in this project, it is again worth mentioning other models of learning where noticing does not play the key causal role. For example, R. Ellis (1994) cites noticing items in input as one of a variety of ways that explicit knowledge can be used in learning, including the possibility that explicitly learned material may be used later at the appropriate stage of development. Also, Swain (1995) cites noticing as one of the benefits of producing output and mentions other examples of output which positively influence learning. Thus, noticing is not necessarily the only factor which is credited with having an impact on learning, and its impact is not always said to be immediate on the learning process. It should also be mentioned in this context the results presented in Chapter 7 concerning individual differences, which pointed to the possibility that good learners are associated with higher noticing scores and proficiency scores. Thus individual differences may have a role to play in developing any equation which suggests a relationship between noticing and learning.

In conclusion, although some of the hypotheses were not confirmed, there are various reasons to opt for an interpretation that does not rule out a possible beneficial role for noticing in input based on the data presented here.

9.3. CONCLUSIONS CONCERNING THE NOTICING TEST ITSELF

As the test for noticing developed for this study was the central data collection instrument used, it is appropriate to briefly discuss and evaluate two aspects of the use of this test.

First, the test does appear to allow for collection of data on different states of consciousness that learners have at the point of encountering input. The audio-recorded retrospective data show that in many cases the detail of experience that participants were able to describe was quite rich. The question arises, though, as to whether participants were able to interpret the content or lack of content of their subjective experience in terms of a Remember, Know, or Guess response — that is, whether they were able to accurately and reliably report the correct response associated with their state of consciousness. As reported above, the few changes that were made in participants' choices during the retrospections point to the fact that participants were relatively accurate in their responses. Furthermore, using the forced-choice

procedure of presenting target items and distractors together as well as the process of adjusting scores for errors leads to a more accurate assessment of these states of consciousness. This type of test, then, would seem to have an advantage over other methods used to assess noticing such as underlining passages in a text (Fotos, 1993) or simply asking participants if they noticed particular features (Robinson, 1995a) by allowing them to provide more detailed responses concerning their state of consciousness. And, as a more detailed and refined use of the concepts relating to consciousness and awareness in SLA studies is needed (Tomlin & Villa, 1994), this noticing test can go some distance in fulfilling that need.

Second, this type of test does require rather extensive training in order for participants to use the instrument correctly. As was described in the methodology section, this training took several forms, including a practice test, the reading and discussion of instructions, and the use of the retrospections. One point to be made is that this training is necessary. But a more important point that needs to be made here is that this training focuses all participants on form at the time of encountering the input. That is, using this method, it would be rather difficult to have a true control group which simply read a text for meaning and then was assessed on their noticing of form. At this point, they would most likely not be able to use the noticing test correctly, or, at the least, the length of the training and discussion that would need to occur at that point would certainly interfere with their memories of what they had just experienced. This inherent focus on form that this testing method brings with it is a large drawback if data are wanted concerning how learners interact with input while focusing on meaning or when interacting in a neutral way.

CHAPTER 10. CONCLUSIONS AND IMPLICATIONS

The purpose of this study has been to explore how instruction and the development of explicit knowledge about a foreign language can lead to a qualitatively different experience with input where aspects of the input are made more salient and therefore consciously noticed, most likely through more elaborative processing in working memory. Next, the second link in the chain was addressed, the connection between noticing and learning. While fairly strong evidence for the impact of instruction on noticing was found, the connection between noticing and learning was shown to be a more tenuous one. Beyond these key results, conclusions can also be drawn about the theoretical approach adopted in this study, and the test of noticing that was developed to measure learners' conscious experience with input.

First, from the results presented here, it appears that the effect that instruction has on noticing is rather short-term and shows a sharp drop-off in strength between a week to ten days after instruction. In the case of vocabulary, it was shown that after this time noticing levels lowered to a point to where the exposure group scored about equally as well on noticing as the instruction group. A possible implication of this is that instructional techniques, tasks, or experiences which are better able to establish strong memories for forms, may be able to extend the period during which these elements are likely to be noticed in input. This should not imply the simple memorization of forms out of context, however, but the active contextualized involvement with forms in a controlled way, which would allow for deeper and more elaborate processing and, hence, the potential for longer retention of those memories.

Second, there is a clear division between the consciousness that learners have of grammar and vocabulary in input. Even though effects for instruction on the noticing of vocabulary were found, scores, in general, were much higher for vocabulary noticing than grammar noticing, both in the instructed and uninstructed conditions. It appears that grammar is simply harder to notice than vocabulary, but unlike with vocabulary instruction where the differences between the instruction and exposure group were evened out at later testing times, stronger results were found for the effects of instruction on the noticing of grammar over time. That is, even though the level of noticing of grammar went down after the first testing time, just like vocabulary, the effect was still large enough at later testing times to show that the instructed group still had an advantage over the exposure group. The implication is that the highlighting of grammatical forms in an instructional setting is a potentially valuable thing to do in that it is accomplishing something that learners may not naturally do on their own. Thus, for forms which have lower saliency, focus on form may have a large payoff in focusing learners' attention to input in a way that might not naturally occur.

While the above conclusion might suggest that vocabulary learning can take care of itself, there are good reasons to believe that vocabulary instruction may be appropriate. N. Ellis (1994) has pointed out that vocabulary learning involves both implicit and explicit processes, where the forms of words can be learned implicitly, and yet where the meanings of words are learned through conscious, explicit processes. Thus, instruction which promotes the conscious noticing of vocabulary in input could play a role in advancing the learning of the meaning of words.

Third, from the data collected from this research project, it appears that even without much prompting, these particular learners are naturally engaging in what sometimes is a quite large amount of noticing of form in input. That is, they appear to be focusing *themselves* on form. Although the reason for this is, without a doubt, due, in a large part, to the heavily form-focused approach that the students are used to, it does raise the question of the degree to which learners in a less academic, or even naturalistic, setting are engaging in the noticing of form for whatever reason.

Fourth, the data has shown that there is considerable variation in individuals in their ability to notice items in input. Furthermore, there were some, though limited, indications in the data that people who are better noticers are better learners. This reflects the possibility that the size of working memory may have a role to play in explaining differences in learning (Harrington & Sawyer, 1992) or that strategies in the way that learners process input may have an effect on noticing and, potentially, learning. Certainly, the data suggest that the issue of individual differences in noticing, whatever their source, is a question to consider.

Finally, concerning the noticing test itself, participants in the study appeared able to distinguish between the states of consciousness of Remembering, Knowing and Guessing. Being able to distinguish between subtle states of awareness of form is important if a clearer conception of the relationship between explicit knowledge and language learning is to be established. Therefore, a test based on the Remember/Know distinction may have future application, especially if this distinction could be used to investigate noticing in more natural, communicative settings. With the continued emphasis on the focus on form in theory and pedagogy, the use of finer-grained tests of memory and consciousness will be important.

APPENDICES

APPENDIX A: AN EXAMPLE OF THE INSTRUCTIONAL MATERIALS USED IN THE STUDY

ROUND ONE, LESSON 1, GRAMMAR INSTRUCTION

Reading

Coincidence: Is it more than just chance?

Sue Hamilton was working alone in her office in July 1992 when the fax machine broke down. After trying unsuccessfully to fix it, she decided to call her colleague Jason Pegler, who had set off home a little earlier. She remembered Jason's writing his number on the notice-board. She found the number, called him up and began to explain the problem. But Jason quickly stopped her. "I'm not at home," he mysteriously explained. What he had done was answer the phone in a phone box that he just happened to be walking by. The number Sue had found on the noticeboard wasn't Jason's phone number. It was his employee number. Amazingly, it turned out to be the same as the number of the phone box that Jason was walking past when she called.

Strange coincidences like this fascinate us, and we humans seem to have a propensity to like these kinds of things. We've all had similar, though perhaps less profound, experiences such as bumping into someone that you know when you're on holiday. There is also the "small world" phenomenon, where, for example, you get your hair cut and find out that you and the barber have a friend in common. Are such experiences merely coincidences or is there some kind of unknown force making these things happen?

Most scientists maintain on principle that coincidences are just the results of the laws of probability. It was a book published in 1898 called *The Wreck of the Titan* which became one of the most famous examples of a coincidence. It depicted the story of the *Titan*, a huge 46,000 ton liner, which its builders claimed was unsinkable. On its maiden voyage from England to New York, it struck an iceberg in the North Atlantic and sank. There were not enough lifeboats on the ship and many of the passengers drowned. Fourteen years later on 15 April 1912, the unsinkable 45,000 ton *Titanic* sank on its maiden voyage from English to New York after hitting an iceberg. Half the passengers drowned because there were not enough lifeboats.

It seems like an strangely correct prediction, but was it? It was probably a succession of lucky guesses condensed into one place. First of all, if you're going to write a book about a ship, isn't it likely that you would choose the biggest ship in the world? And which would be more dramatic, it first voyage or its 23rd, an ordinary ship or an unsinkable ship, everyone survives or there aren't enough lifeboats? Secondly, some facts are the natural results of other choices. A huge liner would probably have a name that means "huge", wouldn't it? And what was a common danger in the North Atlantic? Icebergs, of course. Looking at things in terms

of probability, the strange coincidences don't seem so strange after all, and our human weakness to believe in silly things is betrayed.

Another way of explaining coincidences is that we notice them simply because they are unusual. We are exposed to so many things during the day that only the most unusual or important are noticed. These memories of everyday normal events deteriorate quickly and we don't think about them again. So you remember meeting your neighbor while you were having a walk on an empty beach while on holiday, and you think that's amazing, but you don't remember all the times that you go on holiday without meeting one of your neighbors.

Still many people sustain a belief that a mysterious force must be at work because some coincidences are too hard to explain in terms of probability or selective memory. Take the case of the young architect who in 1971 tried to commit suicide by jumping in front of a London Underground train. The train pulled up just in time and the architect survived. But the train driver hadn't stopped the train. A passenger who had no idea what was happening had deliberately pulled the emergency cord. Talking about the incident later he said that he had suddenly felt "driven" to stop the train and gave the emergency cord a pull. Pure chance or a mysterious force which triggers these events? Can we really be certain?

Comprehension questions: True or false?

1. _____ Jason Pegler's employee number was the same as his phone number.
2. _____ The "small world" phenomenon is when you meet a friend in an unusual place.
3. _____ The details of the wreck of the Titan and the Titanic were the same.
4. _____ The coincidence between the Titan and the Titanic can be explained by probability.
5. _____ There are more coincidences because we have bad memories.
6. _____ The man who pulled the emergency cord didn't really know why he did it.
7. _____ The author of the text thinks there must be a scientific explanation for coincidences.

Talking points

1. Do you agree with the explanations in the article?
2. Can these ideas explain what has happened to you or to your friends?
3. How would you explain the story of the architect?

Grammar exercises

1. Find the underlined structures in the text which are similar in form to the structures below. Write the line number of the structure in the space. (Note: you will probably leave some blank.)

_____ What they did was sit in the shade and sip cool drinks.
_____ He didn't know what they had done.
_____ When they got home, they saw Bob's car in the driveway.

_____ They noticed Jennifer's leaving early from the meeting.
_____ It will be the internet which will revolutionize communication.
_____ It is important to book your tickets early.
_____ They got home after 1 AM.
_____ After he got his car washed at the gas station, he went for a drive.
_____ The woman gave the car a push.
_____ Mary gave the dog a bone.
_____ The farmer had a look around the fields.
_____ Matt had a 1952 Ford pickup.

2. Each of the sentences marked in the text can be written in an alternative way. Write that sentence in the blank provided. What do you think the difference in meaning is (if any)?

Line 3: She remember Jason's writing his number on the noticeboard.

_____

Line 5: What he had done was answer the phone in a phone box that he just happened to be walking by.

_____

Line 12: ...for example, you get your hair cut and find out that you and the barber have a friend in common.

_____

Line 17: It was a book published in 1898 called *The Wreck of the Titan* which became one of the most famous examples of a coincidence.

_____

Line 37: So you remember meeting your neighbor while you were having a walk on an empty beach...

_____

Line 45: Talking about the incident later he said that he had suddenly felt "driven" to stop the train and gave the emergency cord a pull.

_____

3. Finish each of the following sentences in such a way that it means exactly the same thing as the sentence printed above it.

a. He left home and got a job in another city.

What he did was

---

b. She admired that Rachel got such good marks.

She admired Rachel's

---

c. He walked on the beach for a long time.

He had

---

d. Janet listened to the new song.

Janet gave

---

e. My house was painted last week.

I

---

f. Bob was hiding behind the couch.

It was

ROUND ONE, LESSON 1, VOCABULARY INSTRUCTION

Reading

Coincidence: Is it more than just chance?

Sue Hamilton was working alone in her office in July 1992 when the fax machine broke down. After trying unsuccessfully to fix it, she decided to call her colleague Jason Pegler, who had set off home a little earlier. She remembered Jason's writing his number on the noticeboard. She found the number, called him up and began to explain the problem. But Jason quickly stopped her. "I'm not at home," he mysteriously explained. What he had done was answer the phone in a phone box that he just happened to be walking by. The number Sue had found on the noticeboard wasn't Jason's phone number. It was his employee number. Amazingly, it turned out to be the same as the number of the phone box that Jason was walking past when she called.

Strange coincidences like this fascinate us, and we humans seem to have a **propensity** to like these kinds of things. We've all had similar, though perhaps less **profound**, experiences such as bumping into someone that you know when you're on holiday. There is also the "small world" phenomenon, where, for example, you get your hair cut and find out that you and the barber have a friend in common. Are such experiences merely coincidences or is there some kind of unknown force making these things happen?

Most scientists maintain on **principle** that coincidences are just the results of the laws of probability. It was a book published in 1898 called *The Wreck of the Titan* which became one of the most famous examples of a coincidence. It **depicted** the story of the *Titan*, a huge 46,000 ton liner, which its builders claimed was unsinkable. On its maiden voyage from England to New York, it struck an iceberg in the North Atlantic and sank. There were not enough lifeboats on the ship and many of the passengers drowned.

Fourteen years later on 15 April 1912, the unsinkable 45,000 ton *Titanic* sank on its maiden voyage from English to New York after hitting an iceberg. Half the passengers drowned because there were not enough lifeboats.

It seems like an strangely correct prediction, but was it? It was probably a **succession** of lucky guesses **condensed** into one place. First of all, if you're going to write a book about a ship, isn't it likely that you would choose the biggest ship in the world? And which would be more dramatic, it first voyage or its 23rd, an ordinary ship or an unsinkable ship, everyone survives or there aren't enough lifeboats? Secondly, some facts are the natural results of other choices. A huge liner would probably have a name that mean "huge", wouldn't it? And what was a common danger in the North Atlantic? Icebergs, of course. Looking at things in terms of probability, the strange coincidences don't seem so strange after all, and our human weakness to believe in silly things is **betrayed**.

Another way of explaining coincidences is that we notice them simply because they are unusual. We are **exposed to** so many things during the day that only the most unusual or important are noticed. These memories of everyday normal events **deteriorate** quickly and we don't think about them again. So you remember meeting your neighbor while you were having a walk on an empty beach while on holiday, and you think that's amazing, but you don't remember all the times that you go on holiday without meeting one of your neighbors.

Still many people **sustain** a belief that a mysterious force must be at work because some coincidences are too hard to explain in terms of probability or selective memory. Take the case of the young architect who in 1971 tried to commit suicide by jumping in front of a London Underground train. The train pulled up just in time and the architect survived. But the train driver hadn't stopped the train. A passenger who had no idea what was happening had **deliberately** pulled the emergency cord. Talking about the incident later he said that he had suddenly felt "driven" to stop the train and gave the emergency cord a pull. Pure chance or a mysterious force which **triggers** these events? Can we really be certain?

Comprehension questions: True or false?

1. _____ Jason Pegler's employee number was the same as his phone number.
2. _____ The "small world" phenomenon is when you meet a friend in an unusual place.
3. _____ The details of the wreck of the Titan and the Titanic were the same.
4. _____ The coincidence between the Titan and the Titanic can be explained by probability.
5. _____ There are more coincidences because we have bad memories.
6. _____ The man who pulled the emergency cord didn't really know why he did it.
7. _____ The author of the text thinks there must be a scientific explanation for coincidences.

Talking points

1. Do you agree with the explanations in the article?
2. Can these ideas explain what has happened to you or to your friends?
3. How would you explain the story of the architect?

Vocabulary exercises

1. Predicting vocabulary in context
   Examples:
   a. The major points of your plan are clear to me, but the details are still *hazy*.
   b. The king *manifested* his pleasure with a hearty laugh.

2. Find the meaning of the underlined words in the article by using context clues, and then match the word with the definition below.

_____ : to cause something to happen
_____ : affects you very strongly; extreme, great
_____ : done in a planned way, not by accident
_____ : to be introduced to something
_____ : to keep something going for a long period of time
_____ : a general belief or rule

_____ : a number of things one after another
_____ : to make something shorter or smaller, to put something in a smaller space
_____ : a natural habit or trend to behave in a particular way
_____ : to represent or show
_____ : to become worse in condition or quality
_____ : to tell or show something which was a secret; to be disloyal

3. Put one highlighted word from the text in each blank.
   a. Advertisements are often _____ written in bad English.
   b. The problem was how to create and _____ public interest.
   c. After he wrecked his car, he was in a state of _____ disbelief.
   d. The holiday was spoiled by a _____ of rainy days.
   e. The newspaper printed a report which _____ a big debate.
   f. Although he wanted to hide it, his face _____ his grief.
   g. He claimed that the economy worked on a few simple _____ .
   h. His eyesight had begun to _____ as he approached 80.
   i. I tried to _____ the report into as few words as possible.
   j. Squirrels have a _____ to run about very quickly looking for food.
   k. In his first novels the author _____ the quite, simple life of a small town.
   l. The river became very low and _____ the rocks on the bottom.

APPENDIX B: INSTRUCTIONS FOR THE NOTICING TEST

Test instructions

You are now going to take a test to see what words and grammar you remember from the passage you have just read. When you turn the page over, you will see two columns of boxes, one for vocabulary and one for grammar. In each box there are two words or two sentences with particular grammatical structures underlined. The example below shows you what the test will look like:

Example for vocabulary:

0 run	0 Remember
	0 Know
0 walk	0 Guess/don't know for sure

Example for Grammar:

0 He <u>lived</u> there for 10 years.	0 Remember
	0 Know
0 He <u>has lived</u> there for 10 years.	0 Guess/don't know for sure

Only one of each pair was in the text. You will need to carefully look at each pair of items and make two decisions.

First, on the left hand side of the box you see a pair of words or vocabulary. You must decide which of the two items were in the reading passage you just read. Put an X next to the item you think was in the reading.

Second, you need to decide how much you remember about the word or structure which you just put a mark next to, but putting and X next to one of the statements on the right hand side. If you just guessed which one was in the text and you really don't know, then put an X by "guessed/don't know for sure." If you think that the item was really in the reading passage, decide whether or not you actually remember seeing the item in the passage, or whether you are just sure that the item was in the text but you can't remember anything specific about it being there. If you consciously remember something specific about seeing the item in the passage, then check the box next to "remember". If you are sure that the item was there but can't remember the experience of seeing it in the text, then check the box next to "know".

Here are some examples to make the difference between "remembering" and "knowing" clearer. There are many reasons that you might mark something as having "remembered" it. You may remember that when you saw the item in the reading passage there was something interesting about it that made it stand out. For example, it may have stood out because it was a new or unknown word or grammatical structure, and this made you remember it. You may also remember something from the reading passage because you studied it before and it stands out because of that or because the item appeared in the passage in an unusual way that you hadn't seen before. Finally, you may just simply remember the experience of seeing the item

in the text and have a picture in your mind of the word or several words, but can't exactly say what the reason why you remember it. In each case you remember experiencing the word in the text, and so you should put and X next to "remember".

If, on the other hand, you feel very strongly that the item appeared in the reading passage but you can't remember the experience of meeting it in the text, then put an X next to the "know" box to show that you simply know that it was in the passage, but can't recall the specific details of it being there.

There are other examples which can make the difference between remembering and knowing clearer. When you see someone your age walking down the street you may know that this person goes to your school but can't remember anything about the person or how you know them at school. This is an example of knowing. On the other hand, you may see a person on the street who goes to your school and be sure of this because the person's brother or sister is your classmate. In this case you remember the situation of seeing this person with your classmate, and so this is an example of remembering.

You may look back at these instructions at any time during the test, but you may not look back at the reading passage.

APPENDIX C: AND EXAMPLE OF TEXTS USED IN THE NOTICING TEST

ROUND ONE, TEST 1, TEXT FOR NOTICING

Trip to the Kingo Valley

Let me describe the problem that we faced in the summer of 1982. Having reached the highest point of our route according to plan, we discovered a profound problem which the map had not told us about. It was impossible to climb down into the Kingo Valley. The river lay deep between mountain sides that were almost vertical. The trail we had been following had deteriorated to nothing, and we couldn't find any animal tracks, which usually show the best way across country. The slopes were covered so thickly with bushes that we could not see the nature of the ground. We had somehow to sustain our journey and break through to the river, which would give us our direction out of the mountains into the inhabited lowlands.

What our guide did was cut a narrow path through the bushes with his long knife and we followed in quick succession. Then, when we thought we had really reached the river, we found ourselves instead on the edge of a cliff, exposed to high winds and with a straight drop of 1,000 feet to the water below. On principle we knew that it was useless to try to go any further, so we had a look over the side and then climbed back up the hill and began to look for another way down. We climbed, slipped and started a small rock slide, scratched our hands to pieces and finally arrived at the river, although I nearly broke my leg in the process. Happily we strode downhill along its bank without having to cut our way. However, after a few miles the river was compacted into a small, steep-sided gap between rocks and suddenly dropped thirty-five feet over a waterfall. A quick look around betrayed the fact that we were again in trouble: there was no path alongside the river and no way round the waterfall. Because the water was moving so fast, we decided not to give swimming a try.

It was one of the guides with a propensity for solving problems who quickly saw a way of overcoming the difficulty. There was a fallen tree lying upside down over the waterfall with its leafy top resting on the opposite bank below the falls. Without hesitation he deliberately climbed down the slippery trunk to show us how easy it was. Having got to the fork of the tree, he moved hand over hand along a branch for four or five feet with this legs hanging in space, then he dropped onto the flat bank on the other side, throwing his arms in the air like a footballer who has scored a goal, and cheerfully waving us on. We greatly appreciated his showing us the way over the river and eventually to the valley below.

Comprehension questions. Mark whether each statement is either true or false.

_____ 1. The travelers had hoped to get down to the river without much difficulty.
_____ 2. One reason the travelers took so long to get to the river was that it was too hot to move quickly.
_____ 3. To get past the waterfall the guide had to use a fallen tree as a kind of bridge.
_____ 4. The travelers were happy when they reached the river because they knew they were near their destination.

REFERENCES

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259–302). Honolulu: University of Hawai'i Press.

Allen, D. (1992). *Oxford placement test 2*. Oxford: Oxford University Press.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, Mass.: Harvard University Press.

Anderson, J. R. (1995). *Learning and memory: An integrated approach*. New York: John Willey & Sons.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Baddeley, A. (1993). Working memory and conscious awareness. In A. Collins, MA. Conway, S. E. Gathercole & P. E. Morris (Eds.), *Theories of memory* (pp. 163–168). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bialystok, E. (1978). A theoretical model of second language learning. *Language Learning, 28*, 69–84.

Bialystok, E. (1982). On the relationship between knowing and using linguistic forms. *Applied Linguistics, 3*, 181–206.

Carr, T. H., & Curran, T. (1994). Cognitive factors in learning about structured sequences: Applications to syntax. *Studies in Second Language Acquisition, 16*(2), 205–230.

Celce-Murcia, M., & Larsen-Freeman, D. (1983). *The grammar book: An ESL/EFL teacher's course*. Rowley, MA: Newbury House.

Chamot, A. U., & O'Malley, M. J. (1994). Language learner and learning strategies. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 371-392). London: Academic Press.

Chaudron, C. (1985). Intake: On models and methods for discovering learners' processing on input. *Studies in Second Language Acquisition, 7*, 1–14.

Cohen, A. D. (1987). Using verbal reports in research on language learning. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp. 82–95). Clevedon: Multilingual Matters.

Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. M. (1997). Changes in memory and awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General, 126*(4), 393–413.

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics, 5*, 161–170.

DeKeyser, R. M. (1994). How implicit can adult second language learning be? In J. Hulstijn & R. Schimdt (Eds.), *Consciousness in second language learning* (pp. 83–96): *AILA Review*, Vol. 11.

DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 17*(3), 379–410.

DeKeyser, R. M. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom language acquisition* (pp. 42–63). Cambridge: Cambridge University Press.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*(4), 493–533.

DeKeyser, R. (2007). *Practice in a second language*. Cambridge: Cambridge University Press.

Dewhurst, S. A., Graham, J. H., & Barry, C. (1998). Separate effects of word frequency and age of acquisition in recognition and recall. *Journal of Experimental Psychology: Learning Memory and Cognition, 24*(2), 284–298.

Dixon, R. M. W. (1991). *A new approach to English grammar, on semantic principles*. Oxford: Clarendon Press.

Doughty, C. J (1991). Second language instruction does make a difference: Evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition, 13*, 431–469.

Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp 256–310). Oxford: Blackwell Publishing.

Dulay, M., & Burt, H. (1975). Creative construction in second language learning and teaching. In H. Burt & M. Dulay (Eds.), *On TESOL '75* (pp. 21–32). Washington D.C.: TESOL.

Ellis, N. (1994). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. *AILA Review, 14*, 37–56.

Ellis, R. (1990). *Instructed second language acquisition*. Oxford: Blackwell.

Ellis, R. (1993). The structural syllabus and second language acquisition. *TESOL Quarterly, 27*(1), 91–113.

Ellis, R. (1994). A theory of instructed second language acquisition. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 46-79). London: Academic press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. (rev. ed.). Cambridge, Mass: The MIT Press.

Faerch, C., & Kasper, G. (1987). Perspectives on language transfer. *Applied Linguistics, 8*(2), 111–136.

Fotos, S. (1992). Consciousness-raising and noticing through focus on form: Grammar task performance versus formal instruction. *Applied Linguistics, 14*, 385–407.

Fuchs, M., Westheimer, M., & Bonner, M. (1994). *Focus on grammar: An intermediate text for reference and practice*. New York: Addison-Wesley.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16*(4), 309–313.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and non word recognition. *Memory & Cognition, 18*(1), 23–30.

Gardiner, J. M., & Java, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory & Cognition, 19*(6), 617–623.

Gardiner, J. M., & Java, R. I. (1993). Recognising and remembering. In A. F. Collins & S. E. Gathercole (Eds.), *Theories of memory* (pp. 163-188). Hove, England: Lawrence Erlbaum Associates.

Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition, 18*(6), 579–583.

Gardiner, J. M., Kaminska, Z., Dixon, M., & Java, R. I. (1996). Repetition of previously novel melodies sometimes increases both remember and know responses in recognition memory. *Psychonomic Bulletin & Review, 3*, 366–371.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of Remembering, Knowing, and Guessing. *Consciousness and Cognition, 7*, 1–26.

Gardiner, J. M., Richardson-Klavehn, A., & Ramponi, C. (1997). On reporting recollective experience and "direct access to memory systems". *Psychological Science, 8*(5), 391–394.

Gass, S. M (1988). Integrating research areas: a framework for second language studies. *Applied Linguistics, 9*, 198–217.

Gass, S. M. (2003). Input and interaction. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp 224–255). Oxford: Blackwell Publishing.

Gass, S. M., & Selinker, L. (2008). *Second language acquisition* (3rd ed.). Hillsdale, N. J.: Lawrence Erlbaum.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley and Sons.

Hall, J. K., A Cheng, & M. Carlson. (2006). Reconceptualizing multicompetence as a theory of language knowledge. *Applied Linguistics, 27*(2), 220–240.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25–38.

Hashemi, L. (1991). *Cambridge First Certificate Examination practice 2*. New York: Cambridge University Press.

Hatch, E. & A. Lazaraton. (1991). *Design and statistics for applied linguistics*. New York: Newbury House.

Hulstijn, J. H. (1989). Implicit and incidental second language learning: Experiments in the processing of natural and partly artificial input. In H. E. Dechert & M. Raupach (Eds.), *Interlingual processing* (pp. 49–73). Tübigen: Gunter Narr.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 113–125). London: Macmillan.

Hulstijn, J. H., & de Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review, 11*, 97–112.

Hulstijn, J., & Schmidt, R. (1994). Guest editors' introduction. *AILA Review, 11*, 5–10.

Izumi, S., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the output hypothesis: effects of output on noticing and second language acquisition. *Studies in Second Language Acquisition, 21*, 421–452.

Johnson, K. (1996). *Language teaching and skill learning*. Oxford: Blackwell.

Jourdenais, R., Ota, M., Stauffer, S., Boyson, B., & Doughty, C. (1995). Does textual enhancement promote noticing? A think-aloud analysis. In R. Schmidt (Ed.), *Attention and awareness in second language learning* (pp. 183–216). Honolulu: University of Hawai'i.

Kenesei, I., Vágó, R. M., & Fenyvesi, A. (1998). *Hungarian*. London: Routledge.

Krashen, S. 1981. *Second language acquisition and second language learning*. Oxford: Pergamon.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. London: Longman.

Leow, R. (1997). Attention, awareness, and foreign language behavior. *Language Learning, 47* (3), 467–505.

Leow, R. (1998). Toward operationalizing the process of attention in SLA: Evidence for Tomlin and Villa's (1994) fine-grained analysis of attention. *Applied Psycholinguistics 19,* 133–159.

Lightbown, P. (1984). Great expectations: Second-language acquisition research and classroom teaching. *Applied Linguistics, 6*(2), 173–189.

Long, M. H. (1981). Input, interaction and second language acquisition. In H. Winitz (Ed.), *Native language and foreign language acquisition. Annals of the New York Academy of Sciences* (Vol. 379, pp. 259–278).

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston: Heinle & Heinle Publishers.

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.

Országh, L. (1990). *A comprehensive English–Hungarian dictionary* (10th ed.). Budapest: Akadémiai Kiadó.

Peckham, D. W. (2001). Noticing grammar and vocabulary as a goal of instruction. *novELTy Journal, 9*(2), 4–22.

Pienemann, M. (1984). Psychological constrains on the teachability of languages. *Studies in Second Language Acquisition, 6*, 186–214.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology, 39*, 475–543.

Robinson, P. (1995a). Aptitude, awareness, and the fundamental similarity of implicit and explicit second language learning. In R. Schmidt (Ed.), *Attention and awareness in second language learning* (pp. 302–357). Hawai'i: University of Hawai'i Second Language Teaching and Curriculum Center.

Robinson, P. (1995b). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45*(2), 283–331.

Robinson. P. (2001). *Cognition in second language instruction*. Cambridge: Cambridge University Press.

Rosa, E., & O'Neill, M. D. (1999). Explicitness, intake and the issue of awareness. *Studies in Second Language Acquisition, 15*, 511–556.

Schacter, D. L. (1989). On the relation between memory and consciousness: Dissociable interactions and conscious experience. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 355–389). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review, 11*, 11–26.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu: Second Language Teaching & Curriculum Center.

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: a case study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to learn* (pp. 237–336). Rowley, Mass: Newbury House.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*, 367–447.

Sharwood-Smith, M. (1981). Consciousness raising and the second language learner. *Applied Linguistics, 2*, 159–68.

Sharwood-Smith, M. (1991). Speaking to many minds. *Second Language Research, 7*, 118–132.

Sharwood-Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition, 15*, 165–179.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*(1), 34–50.

Soars, J. & L. Soars. (1998). *New headway English course: Upper-intermediate*. Oxford: Oxford University Press.

Strack, F., & Förster, J. (1995). Reporting recollective experiences: Direct access to memory systems? *Psychological Science, 6*(6), 352–358.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics* (pp.125–144). Oxford: Oxford University Press.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive process they generate: A step towards second language learning. *Applied Linguistics, 16*(3), 371–391.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16*(2), 183–203.

Trahey, M., & White, L. (1993). Positive evidence and preemption in the second language classroom. *Studies in Second Language Acquisition, 15*(2), 181–204..

Truscott, J. (1998). Noticing in second language acquisition: A critical review. *Second Language Research, 14*(2), 103–135.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*(1), 1–12.

University of Cambridge Local Examinations Syndicate. (1990). *Cambridge First Certificate Examination practice 3*. New York: Cambridge University Press.

Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing in incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition, 19*(3), 287–308.

White, J. (1998). Getting the learners' attention: A typographical input enhancement study. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 85–113). Cambridge: Cambridge University Press.

Winter, B., & Reber, A. S. (1994). Implicit learning and the acquisition of natural languages. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 115–145). London: Academic Press.

PREVIOUS VOLUMES

1. *Selected Articles by the Members of the English Department.*
   Ed. Annamária Halász (1981).
2. *Selected Articles by the Members of the English Department.*
   Ed. Bálint Rozsnyai (1983).
3. *Shakespeare and the Emblem.*
   Ed. Tibor Fabiny (1984).
4. *Literary Theory and Biblical Hermeneutics.*
   Ed. Tibor Fabiny (1992).
5. Attila Kiss, *The Semiotics of Revenge. Subjectivity and Abjection in English Renaissance Tragedy;* Antónia Szabari, *Demand, Desire and Drive in Sidney's Texts and Their Contexts.*
   Monograph Series 1 (1995).
6. *Proceedings of the Second Conference of HUSSE [Hungarian Society for the Study of English].*
   Ed. György Novák (1995).
7. *Iconography in Cultural Studies (Selected Papers of the Szeged Conference, „Iconography East & West", June 1993).*
   Ed. Attila Kiss (1996).
8. *The Iconography of Power. Ideas and Images of Rulership on the English Renaissance Stage.*
   Ed. György E. Szőnyi & Rowland Wymer (2000).
9. Anikó Németh, *„Art, the Embodied Expression of Man".*
   Monograph Series 2 (2004).
10. *The Iconography of the Fantastic.*
    Ed. György E. Szőnyi, Attila Kiss & Márta Baróti-Gaál (2002, in cooperation with *Studia Poetica*).
11. Irén Annus, *Social Realities in the Making. The Structuration of Society and the Constitution of American Identity.*
    Monograph Series 3 (2005).
12. *Spaces in Transition. Essays in Honor of Sarolta Marinovich-Resch*
    Ed. Erzsébet Barát (2005).
13. Zoltán Vajda, *Innovative Persuasions: Aspects of John C. Calhoun's Political Thought.*
    Monograph Series 4 (2007).
14. Réka M. Cristian & Zoltán Dragon, *Encounters of the Filmic Kind: Guidebook to Film Theories.*
    Monograph Series 5 (2008).
15. *The Iconology of Gender I: Traditions & Historical Perspectives; II: Gendered Representations in Cultural Practices.*
    Ed. Attila Kiss & György E. Szőnyi (2008).

A tördelést a JATEPRINT,
a Bölcsészettudományi Kar Kiadványszerkesztősége végezte
WordPerfect 8 kiadványszerkesztő programmal.

Kiadja a JATEPress
6722 Szeged, Petőfi Sándor sugárút 30–34.
http://www.jate.u-szeged.hu/jatepress/

Felelős kiadó: Dr. Vajda Zoltán egyetemi docens, intézetvezető
Felelős vezető: Szőnyi Etelka kiadói főszerkesztő
Méret: B/5, példányszám: 100, munkaszám: 56/2009.

This book deals with a subject of applied linguistic research that has become popular recently, but has not been explored profoundly enough yet – an area of research that has been discussed in numerous publications in the English-speaking world, but is still not very well-known or examined in Hungary.

In addition to this, the book has a very important characteristic feature: the participants of the research are Hungarian secondary school students, and this fact is not widely known to either Hungarian or international readers as the research as a whole has not been available to a broader audience.

Not only its empirical chapters make this book interesting to read, it also provides a valuable synthesis of cognitive linguistic research through a literature review written in an easy-to-understand way, and with a good critical sense.

The subject area discussed in the present book focuses on one of the basic pillars of theories concerning language acquisition, that is the role of generally accepted language teaching practices in the acquisition of English as a foreign language, the purpose of which practices are to help learners sense and consciously realize new grammatical structures and vocabulary.

The outstanding feature of the empirical research discussed in this book is the carefully planned and accurately implemented study documented in appropriate detail. The results are convincing and reflect the advantages as well as difficulties of classroom research in an excellent way.

The author evaluates the results of his research in a realistic and critical way, showing an excellent example of how to interpret data within the frames of feasibility, and how to set new objectives for further research. The text as a whole is nicely formatted, logically built and well edited. The charts are visually pleasing and support understanding.

This book is very useful for students who are language majors preparing to be teachers both in Hungary and abroad, as well as graduate students and researchers interested in language acquisition research.

Marianne Nikolov PhD, Dss
University of Pécs
Department of English Applied Linguistics