

A LONGITUDINAL STUDY OF THE EFFECTS OF INSTRUCTION ON THE
DEVELOPMENT OF ARTICLE USE BY ADULT JAPANESE ESL LEARNERS

by

JOHN DEAN MELLOW

B.A., The University of Calgary, 1987

M.A., McGill University, 1990

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE

DEGREE OF DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Centre for the Study of Curriculum and Instruction)

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

July 1996

© John Dean Mellow, 1996

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Centre for the Study of Curriculum and Instruction
Department of _____

The University of British Columbia
Vancouver, Canada

Date *July 27/96*

Abstract

This dissertation investigates the effects and value of instructional activities for improving second language use of English articles. After reviewing a number of issues concerning pedagogical, linguistic, psycholinguistic, and internal validity, this study presents the results of eight longitudinal time-series case studies of adult Japanese learners of English residing in Vancouver, Canada, four of whom received grammatical explanations, input processing activities, and output practice activities regarding English article use. Learner development was assessed on three different narrative retelling tasks (spoken, written, and cloze) and the production was analysed with reference to specific contexts of use, indicating the form-function mappings that comprised the learners' interlanguage knowledge. The results indicated that the learners' interlanguage production exhibited (a) the anticipated task variation, with greater supplience of *the* on tasks that allowed greater attention to form, and (b) the anticipated discoursal variation, with *the* supplied more consistently when it was primed as a redundant element on the written task and with *the* supplied less consistently when it was efficiently deleted as a redundant element on the spoken task. The results also indicated the variable nature of individual development and the value of assessing development longitudinally on different tasks. Importantly, the results indicated that the learners improved or continued improving after instruction, and strongly suggested that instruction can cause automatization of interlanguage knowledge. This finding suggests that form-focused instruction may be valuable for second language learning, and that pedagogical positions opposing form-focused instruction may need to be revised or abandoned.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Tables	vi
List of Figures	viii
Acknowledgement	ix
Chapter 1 The Effects of Instruction on the ESL Development of Article Use	1
1.1 Introduction	1
1.2 Theories of Instructed SLA	2
1.3 Empirical Investigations of Instructed SLA	6
1.4 The General Research Questions	11
1.5 A Summary of the Chapters	12
Chapter 2 Second Language Acquisition: Form-function Mappings, Variability and Development Over Time	14
2.1 Alternative Approaches to Linguistic Analysis: Structural and Functional, General Competencies and Interacting Sets of Form-function Mappings	14
2.1.1 Functional Analyses of Language Use and Development	18
2.1.2 Functional Analyses of Article Use and Development	19
2.1.3 Functional Analyses of <i>the</i> Suppliance for Tracking Specific Referent Noun Phrases	23
2.1.3.1 <i>the</i> Use with Respect to the Position of the Noun Phrase Within the Sentence	27
2.1.3.2 <i>the</i> Use with Respect to Referential Distance	28
2.2 Variability in Interlanguage: Task Variation and Individual Variation	29
2.3 A Theory of the Developmental Course of SLA	34
2.3.1 The Temporal Nature of the Developmental Course of Instructed SLA	36
2.4 The L2 Development of Form-function Mappings: Interacting Cognitive and Linguistic Principles	40
2.5 Research Hypotheses	53
Chapter 3 Issues of Research Design and Measurement in Instructional Studies	59
3.1 A Methodological Review of Instructional Studies	59
3.2 Experimental Design Issues	64
3.3 The Times-series Design	67
3.3.1 Time-series Designs: Internal Validity	70
3.3.2 Time-series Designs: External Validity	72
3.3.3 Time-series Designs: Measurement issues	75
3.3.3.1 Establishing the Validity of Instruments	76
3.3.3.2 Establishing the Reliability of Instruments	76
3.3.3.3 Establishing Appropriate Measurement Scales	81

Chapter 4	Method	84
4.1	The Method of the Current Study	84
4.1.1	Learners	85
4.1.2	Design	88
4.1.3	The Data Collection Procedure	90
4.1.4	Measurement Instruments	92
4.1.5	Instruction	100
4.2	Data Analysis	107
4.2.1	Transcription and Coding	107
4.2.2	Analysis of the Data: Variables and Summary Measures	112
4.2.2.1	Definitions of the Five Dependent Variables	112
4.2.2.2	Descriptions of the Four Summary Analyses	114
4.2.2.2.1	Group Mean Phase Scores	115
4.2.2.2.2	Inferential Statistical Tests	116
4.2.2.2.3	Phase Regression Lines	119
4.2.2.2.4	Line Graphs	122
Chapter 5	Results and Discussion	123
5.1.	PreI Phase Indications of Interlanguage Patterns	123
5.1.1	Hypothesis 1: Task Variation, PreI Phase	123
5.1.2	Variation According to Linguistic Context during the PreI Phase	126
5.1.2.1	Hypothesis 2: Syntactic Position, Spoken Task	126
5.1.2.2	Hypothesis 3: Syntactic Position, Written Task	127
5.1.2.3	Hypothesis 4: Referential Distance, Spoken Task	130
5.1.2.4	Hypothesis 5: Referential Distance, Written Task	132
5.1.3	An Examination of the Equivalence of the Groups	134
5.2.	The Effects of Instruction on the Five Dependent Variables	138
5.2.1	Hypothesis 6: Group Differences, Group Mean Phase Scores, All Dependent Variables, All Tasks	138
5.2.2	PreI - PostI Changes in the Regression Lines	145
5.2.2.1	Hypothesis 7a: Treatment Group, Gains, Regression Lines, All Dependent Variables, All Tasks	150
5.2.2.2	Hypothesis 7b: Group Differences, Regression Lines, All Dependent Variables, All Tasks	151
5.2.3	Variation According to Linguistic Context and Task, PreI - Post I Changes, Group Mean Gain Scores and Regression Lines	153
5.2.3.1	Hypothesis 8a: Syntactic Position, Spoken Task, Treatment Group, Changes	153
5.2.3.2	Hypothesis 8b: Syntactic Position, Spoken Task, Group Differences, Changes	154
5.2.3.3	Hypothesis 9: Syntactic Position, Written Task, Both Groups, Changes	157
5.2.3.4	Hypothesis 10: Referential Distance, Spoken Task, Both Groups, Changes	159
5.2.3.5	Hypothesis 11: Referential Distance, Written Task, Both Groups, Changes	163
5.3	The Developmental Course of Instructed SLA: An Examination of Individual Development	166
5.3.1	Hypothesis 12: Developmental Course, Individual Patterns	166
Chapter 6	Limitations, Summary, and Implications	173
6.1	Limitations	173
6.2	Summary of Findings	175
6.3	Implications and Future Directions	179

References	182	
Appendix A	Consent Form	197
Appendix B	Line Graphs for All Learners for All Dependent Variables	198
Appendix C	The 11 Aesop's Fables Used in the Study	207
Appendix D	Time Sequence Graphic, Example for One Fable	209
Appendix E	Composition Worksheet, Example from One Learner (Tr4)	210
Appendix F	Cloze Worksheet, Example from One Fable	211
Appendix G	Article Instruction, Grammar Presentation Sheet	212
Appendix H	Scoring Chart, Example from One Learner (Tr4)	214
Appendix I	Results for All Learners, All Dependent Variables, Means	215
Appendix J	Group Regression Lines for All Dependent Variables	217

List of Tables

Table 2.1	Bickerton's Four Categories of Article Use, from Huebner (1979), with Examples Based upon Tarone and Parrish (1988, p. 27)	21
Table 2.2	Hypothetical Development Stages in the Acquisition of Article Use with Referentially Definite Noun Phrases	22
Table 2.3	Possible Interlanguage Form-function Mappings of Articles	42
Table 3.1	Number of Instructional Studies Exhibiting a Positive Effect or No Effect Due to Instruction, According to Delay Until Posttest and Type of Test	62
Table 4.1	Order of Counterbalancing of Different Fable Versions	94
Table 4.2	Four Parameters of the Three Retelling Tasks	97
Table 4.3	Total Properties of Tasks that Contribute to Attention to Form and Use of Rules	99
Table 5.1	Percentage of <i>the</i> Suppliance for Referential Definites on the Three Different Tasks During the PreI Phase	124
Table 5.2	Slopes of the Regression Lines for the Spoken and Written Tasks During the PreI Phase for All Five Dependent Variables	125
Table 5.3	Group Mean Percentage of <i>the</i> Suppliance in Subject and Non-Subject Positions on the Spoken and Written Tasks during the PreI Phase	126
Table 5.4	Predicted Scores for Percentage of <i>the</i> Suppliance on Day 1 and Day 80 in Subject and Non-Subject Positions on the Written Task During the PreI Phase	129
Table 5.5	Percentage of <i>the</i> Suppliance for Current and Recent NPs on the Spoken and Written Tasks During PreI Phase	131
Table 5.6	Slopes of the Regression Lines for Treatment and Control Groups, Spoken Task, PreI Phase	135
Table 5.7	Patterns in Percentage of <i>the</i> Suppliance for Treatment and Control Groups According to Linguistic Context, PreI Phase, Assessed with Group Mean Phase Scores	135
Table 5.8	Patterns in Slope Values of <i>the</i> Suppliance for Treatment and Control Groups According to Linguistic Context, PreI Phase	136
Table 5.9	Slopes of the Regression Lines for Treatment and Control Groups, Written Task, PreI Phase	137
Table 5.10	Patterns in Slope Values of <i>the</i> Suppliance for Treatment and Control Groups According to Linguistic Context, PreI Phase	137
Table 5.11	Gain Scores, Based on the Group Mean Percentages of the Suppliance on the Three Different Tasks for Each of Five Dependent Variables	139
Table 5.12	Differences between Group Gains in Mean Percentage of <i>the</i> Suppliance (i.e., TrGgain - ConGgain) on the Spoken and Written Tasks for Each of the Five Dependent Variables	140
Table 5.13	Slope Values and Estimated Scores on the First Day of the PostI Phase of Percentage of <i>the</i> Suppliance for the Treatment and Control Groups on the Spoken and Written Tasks for Each of Five Dependent Variables	146
Table 5.14	Categories of Changes in Regression Lines from the PreI Phase to the PostI Phase	147
Table 5.15	Mean Phase Scores for the Treatment Group for <i>the</i> Suppliance in Subject and Non-Subject Positions on the Spoken Task	154
Table 5.16	Mean Phase Scores for the Control Group for <i>the</i> Suppliance in Subject and Non-Subject Positions on the Spoken Task	155
Table 5.17	Estimated Day 80 Scores (from the PreI Regression Lines) and Mean PostI Phase Scores for the Treatment Group for <i>the</i> Suppliance in Subject and Non-Subject Positions on the Written Task	158
Table 5.18	Estimated Day 80 Scores (from the PreI regression lines) and Mean PostI Phase Scores for the Control Group for <i>the</i> Suppliance in Subject and Non-Subject Positions on the Written Task	159

Table 5.19	Mean Phase Scores for the Treatment Group for <i>the</i> Suppliance for Recent and Current Noun Phrases on the Spoken Task	160
Table 5.20	Mean Phase Scores for the Control Group for <i>the</i> Suppliance for Recent and Current Noun Phrases on the Spoken Task	160
Table 5.21	Mean Phase Scores for the Treatment Group for <i>the</i> Suppliance for Recent and Current Noun Phrases on the Written Task	163
Table 5.22	Mean Phase Scores for the Control Group for <i>the</i> Suppliance for Recent and Current Noun Phrases on the Written Task	164
Table 6.1	Results of Hypotheses 1 Through 5	175
Table 6.2	Results of Hypotheses 6 Through 11	176
Table 6.3	Principles Regarding the Interaction of Three Factors that Affect Variation in L2 Use	180

List of Figures

<i>Figure 2.1</i>	Possible temporal learning curves	37
<i>Figure 4.1</i>	Sample activity of affective input processing instruction, excerpted from the second exercise	104

ACKNOWLEDGEMENT

This dissertation would not have been completed without the assistance of many, many people. First, I would like to thank my committee members: my supervisor, Ken Reeder, whose guidance has gone well beyond the domain of dissertation writing, into scholarship, academia, and life; and my other committee members, in alphabetical order: Rick Berwick, especially for help in data collection and framing psychometric issues; Alister Cumming, for constant advice, feedback, expertise, support, and mentoring that extends far beyond this document; Patsy Duff, for invaluable help with clear critical thinking and conceptual issues; and Joerge Roche, who has been instrumental in shaping my view of language.

I would also like to thank the students who participated in this study: for their commitment to this project, for teaching me about language learning, and for making eight months of data collection enjoyable and interesting. I am very grateful for the assistance that others have provided me: for help with transcription, coding, and data analysis, Sandie Kouritzin Nakagawa and Anne O'Sullivan; for statistical advice, Holmes Finch; for help with data collection, Jean Hamilton, Cheri MacLeod, and Bill McMichael; for help with printing this final version, Keith McPherson and LERC.

There are many other people to whom I am indebted for discussions, for copies of their papers and dissertations and instruments, and for comments and ideas that have helped me clarify some of my own ideas. An incomplete list of these people includes: Craig Chaudron, Margaret Early, Phillip Elliot, Doug Flahive, Janet Fuller, Robert Hamilton, Ho Han, Mike Long, Peter Master, Deborah Masterson, Bernie Mohan, Teresa Pica, Peter Robinson, Larry Selinker, Karen Stanley, Elaine Tarone, Bill VanPatten, Lydia White, and Kate Wolfe-Quintero.

I am also very grateful for the funding that I received while I worked on this project: a doctoral fellowship from the Social Sciences and Humanities Research Council of Canada (#753-91-0290) and Sir James Lougheed Award of Distinction from the Province of Alberta.

There are many others to thank for support. Another incomplete list includes my family, Dave, Joelle, Margaret, Marlene, and Megan, for their constant support; to Gerald and John, especially for help with data analysis and fruit salad; to John Swenson, especially for software support and advice on high mileage; to Carol, for psychotherapy, and to Ed, especially for the day in Maple Ridge; to Gerald, for the typing; to Roger, especially for the four hour mountain run; to Cynthia, Dave, Ed, Leigh, Wendy, Leslie, Juliet, and Janet for their friendship and support; and to Julius Erving, whose namesake this will finally make me.

This dissertation is dedicated to my parents, Rodger and Carol, for the long-term effects of their instruction.

Chapter 1

The Effects of Instruction on the ESL Development of Article Use

1.1 Introduction

The study in this dissertation examines the development of English article use by second language (L2) learners. The study is concerned with utterances produced by English as a second language (ESL) learners that are similar to the sentence in 1) below. In contrast, a native speaker of English is expected to produce the utterance in 2).

- 1) Woman is talking to girl
- 2) The woman is talking to the girl.

In 1), *woman* and *girl* are bare nouns: they are not preceded by articles. In 2), the target-like sentence, each noun is preceded by an article, *the* in this example, given the assumption that the two noun phrases are definite. Although native English speakers always acquire the knowledge and ability to use articles, ESL learners, even those who have had language classes and are quite proficient in other aspects of English, have great difficulty with this element, often omitting it when it would be supplied by native speakers of English (e.g., Celce-Murcia & Larsen-Freeman, 1983; Master, 1987). Although other aspects of the discourse or context may provide the meaning conveyed by the article, the omission or incorrect use of articles may clearly mark a speaker as non-native and may interfere with or delay comprehension by listeners. For this reason, this study investigates the ways that L2 learners use articles and whether certain instructional activities may help learners to develop toward more target-like use. The study therefore investigates the effect of instruction on the nature of *interlanguage*, the systematic linguistic knowledge that develops along a continuum from the first language to the L2 (cf. Larsen-Freeman & Long, 1991, p. 60; Selinker, 1972).

The value of teaching grammatical items, such as articles, has been critically considered by many language educators and researchers (e.g., Collier, 1992; Ellis, 1990a; Hammerly, 1992; Harley, Cummins, Swain, & Allen, 1990; Higgs & Clifford, 1982; Krashen, 1982, 1992, 1993;

Lightbown & Spada, 1993; Long, 1983, 1988; Swain, 1985; VanPatten, 1993). Because there is considerable disagreement among teachers and researchers as to whether or how grammar should be taught, and because of the difficulty in attaining target-like mastery in a second language, the purpose of this dissertation is not to resolve this issue. Given the complexity of second language acquisition (SLA), one would not expect that any specific study, or even any set of studies, will reveal a simple pedagogical solution. Instead, the dissertation seeks to address the underlying assumptions and design limitations of research into instructional outcomes, to outline linguistic and cognitive processes that may account for L2 use and development, and then to report the results of a study of these issues. More specifically, the study investigates (a) whether the article *the* will be supplied more often on tasks that permit or encourage greater attention to linguistic form; (b) whether the interaction of a series of linguistic and cognitive principles can account for certain aspects of *the* suppliance on different tasks and in different linguistic contexts; (c) whether instruction will result in improvements in *the* suppliance, with such improvements on a spoken task indicating automatization of L2 knowledge; and (d) whether instructed L2 development is characterized by specific developmental patterns, such as U-shaped development and gradual development. The results of this study therefore indicate important directions for theory, pedagogy, and research in second language learning (cf. Gregg, 1989, 1993; Long, 1993; Mellow, Reeder & Forster, in press). Several important issues regarding theories and empirical investigations of instructed second language acquisition are introduced in the next two sections.

1.2 Theories of Instructed SLA

A large number of theories have been proposed to explain how L2 learning occurs (cf. Ellis, 1994; Gass & Selinker, 1994; Larsen-Freeman & Long, 1991; Lightbown & Spada, 1993). Some of these have been derived from specific theories of language, such as structuralist, generative, or functionalist approaches to linguistics. Other accounts of L2 learning have been derived from various theories of learning, such as behaviorist, parameter-setting, or cognitive theories. Many of these theories have specific implications for pedagogy. For the purpose of this

introduction, the many different theories of instructed SLA will be considered with respect to three sets of criteria regarding the nature of language learning: 1) emergence and construction, 2) input and output, and 3) form and function. Although these categories are used here to evaluate approaches to SLA, they could also be used to distinguish approaches to first language acquisition, with the determinants and nature of first language acquisition potentially differing from those of SLA.

The first category for distinguishing between approaches to SLA refers to the degree to which a learner employs active and extensive cognitive processes that use extensive language input to construct linguistic knowledge. The two ends of the continuum in this category are described as *emergence* and *construction*¹. The emergence view suggests that (a) language emerges in a learner according to the learner's own internal syllabus, largely as a result of innate, biological, language-specific predispositions; (b) language development results only partially from the learner's general cognitive operations (e.g., syntactic rules are not learned with processes such as generalization or unconscious hypothesis testing; or, deliberate, deductive practice and exercises contribute only minimally to development); and (c) language development relies only partially on the linguistic environment (e.g., the environment does not provide sufficient information for a learner to construct a knowledge of language) or responds only to certain types of linguistic environments (e.g., instruction and correction of forms do *not* contribute to development). The alternative view, construction, minimizes the importance of innate, language specific abilities, and suggests that language develops largely as a result of extensive cognitive processes and in response to language input (including instruction and correction), output (including practice of various types), and interaction. This categorization of SLA theories is particularly important for educators because an adherence to a viewpoint at either end of the continuum results in specific claims about the types of instruction that should be used in classrooms.

¹ The two categories, emergence and construction, are similar in certain respects to the categories of determinants of language acquisition proposed by other researchers, including Larsen-Freeman and Long (1991, pp. 227-287): nativist, environmentalist, and interactionist; and Ingram (1989, pp. 25-9, 69-73), for L1 acquisition: behaviorist, nativist-maturational, and nativist-constructivist.

Adopting assumptions similar to those of a principles and parameters approach to language and language learning (e.g., Chomsky, 1981; Chomsky & Lasnik, 1993; White, 1989; cf. Mellow, 1988), the very influential view of Krashen (1982, 1992, 1993) assumed an emergent view of language learning and suggested that form-focused instruction is therefore of relatively little value (for further discussion, see Gregg, 1984; Larsen-Freeman & Long, 1991; among others). Breen (1987) discussed pedagogical approaches that emphasize the learner's internal syllabus rather than any external syllabus imposed through form-focused instruction. Other approaches to language pedagogy, including many that are labeled with the term *communicative*, also assume aspects of an emergent approach. In these views, language development results from only a limited subset of the learner's cognitive operations (e.g., deliberate, deductive practice of predetermined linguistic items contributes only minimally to development), and from only a limited type of environmental input types (e.g., input, interaction and communication are crucial, but production practice, instruction, and correction of forms contribute minimally to development) (e.g., Cummins, 1987; Goodman et al., 1987, pp. 161-177, regarding L1 pedagogy; Graman, 1988; cf. Ellis, 1994, pp. 652-654).

In contrast, other approaches to L2 pedagogy adopt a construction view of language learning. Theories of L2 pedagogy such as Grammar-translation and Audio-lingualism stress learner construction of language ability through metalinguistic knowledge and practice (e.g., Grammar-translation), or habit-formation (e.g., Audio-lingualism) (cf. Larsen-Freeman, 1986; Richards & Rodgers, 1986). Those approaches to L2 pedagogy that utilize the drill and practice of language structures also adopt a construction approach (e.g., Hammerly, 1992; Hubbard, Jones, Thornton, & Wheeler, 1983). In addition, a construction approach may emphasize the value of consciousness-raising or input enhancement regarding certain aspects of the language being learned (e.g., Sharwood Smith 1981, 1991, 1993; White, Spada, Lightbown, & Ranta, 1991).

Although not denying the possibility of biological constraints on the form or processing of language nor the importance of communicative interaction, the position that will be investigated in this dissertation is that construction of language knowledge by the learner plays a significant role in

the developmental process. In particular, this view of learning is informed by cognitive theories that include processes of automatization and restructuring (cf. McLaughlin, 1987, pp. 133-153, 1990; Ellis, 1990a, pp. 175-184; Bialystok, 1988), in which automaticity increases over time and as a result of practice. As a result, the type of instruction investigated in this study includes practice activities involving article use. In addition, the processes of automatization and restructuring play a central role in this study's analyses of the development of the ESL use and development of articles.

The second categorization of approaches to SLA distinguishes between the role of comprehension (or input) and the role of production (or output). The nature of both input and output processes have been carefully investigated in recent studies (for overviews, see Ellis, 1994; Gass & Selinker, 1994; Larsen-Freeman & Long, 1991). A number of researchers have discussed and investigated the importance of interaction and negotiation for providing a learner with comprehensible input (e.g., Pica et al, 1993, among many others). In addition, other researchers have clarified the nature of the processing of input, attempting to distinguish types or stages of input processing, utilizing concepts such as intake, incomprehensible input, noticing, apperception, salience, and consciousness (e.g., Cadierno, 1992; Corder, 1967; Gass & Selinker, 1994; Schmidt, 1990; Schmidt & Frota, 1986; Terrell, 1991; VanPatten, 1993; White, 1987). Likewise, the relevance of output for language learning has also been investigated by researchers such as Swain (1985, 1993) and Cumming (1990). Some approaches to language teaching emphasize the role of output practice (e.g., Grammar-Translation), while other approaches have suggested that only input will cause changes in a learner's internalized knowledge (e.g., Krashen, 1982; cf. VanPatten, 1993). The position that will be investigated in this dissertation is that both input and output processes are important for development and, therefore, that optimal instructional activities should involve both comprehension and production, and especially should insure that input is noticed and comprehended.

The third categorization of approaches to SLA distinguishes between a focus on form (or structure) and a focus on function (or meaning). A focus on form or structure has characterized a

number of approaches to SLA theory (e.g., Universal Grammar: White, 1989) and pedagogy (e.g., Grammar-translation, Audio-lingualism). In contrast, a focus on function or meaning has characterized other approaches to SLA theory (e.g., Pfaff, 1987; Tomlin, 1990) and pedagogy (e.g., van Ek, 1987; Graman, 1988). Observing this dichotomy between form and function, a number of researchers have emphasized the importance for the learner of the simultaneous processing of both the structure and meaning of language, resulting in mappings between form and function. Cumming (1990) discusses the value of concurrent metalinguistic and ideational thinking, and Terrell (1991, p. 56) discusses the linking or binding of meaning and form. Thus, an optimal learning event may be achieved by attending to form and meaning simultaneously, rather than just form or just meaning (cf. Brown, 1987, pp. 65-70; Cadierno, 1992; VanPatten, 1993; 1994). The position that will be investigated in this dissertation is that a consideration of form-function mappings is essential for understanding the developmental process and that learning activities should involve, if possible, a concurrent focus on form and meaning.

1.3 Empirical Investigations of Instructed SLA

Given the diversity of approaches to SLA theory and pedagogy, empirical investigations have often been undertaken in order to assess the validity of different theoretical and pedagogical positions (for overviews, see Cadierno, 1992; Chaudron, 1988; Ellis, 1990a, 1994; Larsen-Freeman & Long, 1991; Long, 1983, 1988). There are a number of important, almost axiomatic, assumptions which underlie such research into instructional outcomes. In this section, five important assumptions regarding empirical investigations are considered.

The first assumption is that there are truths regarding instruction that can be determined: (a) that for all or many learners, and for all or many language elements, certain instructional activities can promote mastery of a second language; (b) that there are certain instructional activities that are more effective than others; and (c) that researchers can actually determine which instructional activities are best or most effective (for discussions of traditional, scientific, ends-means approaches to pedagogy, see Hangsen, 1991; McKenna, 1976; Tyler, 1949). This is an

important and potentially controversial assumption: one might not assume that research could prove that there are *best* methods for aiding other aspects of human cognitive development, such as creative artistic expression, critical thinking skills, or the appreciation of music. This assumption also implies that second language acquisition is not too variable nor too complex. Instructed SLA is a complex phenomenon, influenced by the learning environment, including the classroom and teacher, and by many learner variables, including a learner's age, first language, educational history, language opportunities outside of the classroom, personality, learning style, motivations, etc. It may be that this variability and complexity precludes clear and definitive generalizations (cf. Gregg, 1984, p. 95; Larsen-Freeman & Long, 1991, p. 19). Although only a small number of learners and a small number of factors are considered in this study, which is insufficient for the determination and verification of such truths, the research program presented here does assume that, in principle, generalizations are possible and truths may be established.

In addition to the assumption that informative and generalizable research results are possible, a second related assumption is that *researchers* should assess the value of instruction. This assumption implies that researchers may be more capable than teachers of assessing whether instruction actually changes learners' behavior in the desired manner. Because this calls into question a teacher's abilities, it is important to clarify such an assumption. A more neutral means of addressing this issue is to ask why it may be difficult for teachers to accurately evaluate the effects of instructional activities. The answer to this may be that language use and development is so complex and variable that progress is difficult to assess. For example, as a result of U-shaped development (e.g., temporary over-generalizations by learners), errors (or non-target-like use) may actually be an indication of increasing competence with a language. In addition, as a result of learners' avoidance of difficult items, the absence of errors is not necessarily an indication of competence with language. And, because learners may perform perfectly on a paper and pencil grammar test, but may make errors when talking outside the classroom to friends, using language in a classroom context does not mean that a learner is able to use that knowledge in all contexts. In response to this complexity, the research in this dissertation, along with considerable other

research in SLA, explores whether important findings may be possible if research questions are proposed and investigated in a careful and theoretically informed manner. In sum, the research in this dissertation assumes that there are regularities in how humans respond to language instruction, that these regularities are difficult to determine, and that a careful and principled study of language development by researchers can yield insights which may not be apparent to a teacher in the complex and demanding context of providing regular language instruction to a number of language learners.

A third important assumption of instructional studies is that they are linguistically and psycholinguistically valid. Specifically, a detailed understanding of the nature of interlanguage use and development is necessary in order to assess how instruction affects that development. This concern was raised by Larsen-Freeman and Long (1991, p. 240), who suggested that a number of approaches to and analyses of interlanguage (IL) development are not adequately based on a theory of language learning and therefore suffer from a "static, target-language orientation." This assumption relates to the type of language analysis that informs a study of the effects of instruction. For example, a study of the effects of instruction might reveal that instructed learners became 20% more accurate (or target-like) in producing a specific grammatical morpheme or structure, such as English articles, in all contexts. However, an instructional study can also or alternatively consider how changes in interlanguage knowledge have contributed to the manner in which the learner expresses or packages intended meanings, and how this new knowledge may be used differently in different linguistic contexts (e.g., Huebner, 1979, 1983a, 1983b) and on different tasks (e.g., Mellow & Cumming, 1994). Importantly, this type of analysis might indicate why overall accuracy does not improve as significantly as expected, perhaps because of U-shaped development or because changes occur to specific uses of a form without causing substantial changes in overall use. By using an analysis that focuses on interlanguage rather than on the target language, this dissertation will attempt to clarify the effects of instruction on SLA. The issue of linguistic and psycholinguistic validity is examined in detail in Chapter 2.

A fourth important assumption of instructional studies is that they have been designed so that the results are pedagogically relevant. Chapter 3 presents a review of studies that evaluated the effects of teaching specific morpho-syntactic elements. The review suggests that instruction almost always has beneficial effects, but that the beneficial effects that have been measured may not be particularly relevant to the needs of teachers and learners (cf. Krashen, 1992). Thus, these studies have indicated that instruction almost always results in significant short term improvement if that improvement is measured on instruments for which students have time to draw upon their conscious metalinguistic knowledge. Because educators may be interested in long term improvement in spontaneous language use, it is important that instructional studies are carefully designed so that they may investigate a construct described as the *developmental course of SLA* (cf. Mellow, Reeder, & Forster, in press). This implies that research must consider whether the effects of instruction are delayed, gradual, and lasting, and whether that linguistic knowledge can become automatic. In order to answer these questions, learner language must be elicited on both monitored and unmonitored tasks and many data samples must be longitudinally collected in order to understand how language use may change over time. The nature of the developmental course of SLA will be discussed in further detail in Chapter 2 and the issue of pedagogical relevance is examined in Chapter 3.

In addition to being pedagogically relevant, a fifth assumption of instructional studies is that they will be of pedagogical value: that the results will actually influence pedagogy. One part of this assumption implies that the results can be presented and interpreted in such a way as to influence the decisions made by teachers and curriculum designers. This implies that language educators may be influenced by experimental results which are statistically significant (e.g., $p < .05$) or by detailed descriptions of the effects of specific methods on specific students (cf. Harley, 1989a). If, on the other hand, teachers are skeptical of the value and results of research, such critical views may seem justified by the succession of researchers who have reported and even preached that research has proved that method X is most effective and that method Y is not effective (cf. Delpit 1986, 1988). Therefore, it is important to acknowledge that education is both

a political and personal process. Teaching may change as a result of various political or ideological pressures and as a result of the various personal beliefs of a teacher, rather than due to empirical or theoretical implications and prescriptions. Teaching will generally be based upon the goals, procedures, and context set forth by employers (who may or may not be the students), as interpreted by individual teachers, in response to their theories and assumptions (implicit or explicit) about language teaching, and in response to their evaluations of the students' needs, abilities, and backgrounds.

This fifth assumption may be the most problematic. To address this problem, the implications of this study are presented in a general theoretical manner that may relate directly to educators' theories of language learning. Therefore, two general theoretical positions regarding instructed SLA are investigated in this study, positions that have already been referred to as construction and emergence. The construction view of language learning, emphasizing the benefits of practicing language elements, is effectively represented by McLaughlin (1990, pp. 125, 126):

In this view, practice can have two very different effects. It can lead to improvement in performance as sub-skills become automatized, but it is also possible for increased practice to lead to restructuring and attendant decrements in performance as learners reorganize their internal representational framework. . . . Performance may follow a U-shaped curve, declining as more complex internal representations replace less complex ones, and increasing again as skill becomes expertise.

In contrast, the view of language learning as emergence, emphasizing the value of comprehension and use, and arguing against the beneficial effects of form-focused instruction, is represented by Krashen (1982):

The acquisition-learning distinction is perhaps the most fundamental of all the hypotheses to be presented here. It states that adults have two distinct and independent ways of developing competence in a second language. . . .this process of converting learned rules into acquired rules was called 'internalization'. Despite our feelings that internalization does occur, the theory predicts that it does not, except in a trivial way. Language acquisition, according to the theory presented in Chapter II, happens in one way, when the acquirer understands input containing a structure that the acquirer is 'due' to acquire, a structure at his or her "i + 1". (pp. 10, 84)

A similar emergent viewpoint is expressed by Graman (1988): "[A]mple research in second language learning indicates that people learn language best when the focus is on content, as in

natural language acquisition, rather than on the explicit instruction of grammar, vocabulary, and pronunciation” (pp. 438-439).

The study in this dissertation investigates the validity of the construction view of language learning. Therefore the pedagogical value of this study may be found in the extent to which the results either support existing beliefs that language learning is construction or result in a reassessment of the assumption that language learning is only or primarily emergence. If, following Johnson (1989, p. 1), curriculum is defined as “all the relevant decision making processes of all the participants”, then this and other studies of instructional effects may influence second language curricula if the results enrich the views of language learning that inform the decision making processes of the teachers, administrators, curriculum writers and other participants involved in second language education.

Given this discussion of these five assumptions, I assume that there is a role for instructional research that seeks to establish causality. In other words, I assume that L2 pedagogy will benefit from experimental or quasi-experimental research that attempts to validly determine the effects of intervention or treatment. Within the field of instructed SLA, the results of quasi-experimental studies may complement the results of other research methods, including descriptive case studies of language learning and studies of the knowledge of expert teachers.

1.4 The General Research Questions

The issues introduced above indicate the scope and significance of the questions that are studied in this dissertation, an investigation of the effects of specific sessions of grammatical instruction on the development of specific aspects of English article use by adult speakers of Japanese. In particular, three issues raised in the third and fourth assumptions provide the central motivation for the research questions of this dissertation and have influenced the choice of the design and measurement instruments. The three issues regarding instructed development are functional variation, task variation, and temporal variation. These three issues lead to the following general research questions, listed in 3), 4), and 5) below.

3) How does instruction affect language use, not only in terms of overall target-like use, but also in terms of the use of specific language elements within functional interlanguage systems, as indicated through use in specified linguistic contexts? Specifically, within the context of the tracking of referents, what functions does the article *the* serve in a learner's interlanguage and how are those functions affected by instruction?

4) How does instruction affect language use on different tasks? Specifically, is *the* omitted more often on spoken tasks that allow little time and cognitive resources for attention to forms (planning and monitoring) and produced more consistently in written tasks that allow greater attention to forms (cf. Tarone & Parrish, 1988), and if so, are these differences affected by instruction? In addition, is there an interaction between task and linguistic context, such that redundant elements are primed in writing but efficiently deleted in speech (Mellow & Cumming, 1994), and if so, are these differences affected by instruction? Do the answers to these two task-oriented questions provide evidence that instruction leads to the automatization of linguistic knowledge?

5) How does instruction affect language use over time? Specifically, are the effects of instruction temporary or lasting (long-term or mid-term), and immediate, delayed, gradual, or U-shaped (cf. Hedgcock, 1991)? Will temporal changes in learning be reflected differently on different tasks?

1.5 A Summary of the Chapters

The dissertation is organized in the following manner. Chapter 1 has indicated the scope of the problem that is investigated, especially in reference to pedagogical issues. Chapter 2 examines the nature of interlanguage form-function mappings in detail, considering functional interlanguage systems, the variability of second language use, and temporal changes in those systems and that variability. The chapter concludes with a set of six principles that integrate these issues and result in a series of twelve research hypotheses.

Chapter 3 examines the methodological challenges involved in establishing a causal relationship between sessions of instruction and changes in interlanguage systems. The chapter begins with a review of previous instructional studies, focusing on methodological issues, and then justifies the methodological solutions adopted in this dissertation. Chapter 4 describes the procedures used in the study, including descriptions of the learners in the treatment and control groups, the design, the data collection procedure, the measurement instruments, and the instruction. Chapter 4 also describes the nature of the data analysis, including descriptions of the

transcription and coding, definitions of dependent variables, and explanations of the statistical analyses that were chosen to summarize the group data in terms of central tendency and trend.

Chapter 5 presents and discusses the results of the study. The chapter begins with a discussion of the preintervention phase behavior of the groups, in terms of linguistic and task variation, and in terms of the equivalence of the two groups. The second section of Chapter 5 provides a detailed discussion of the effects of instruction on *the* suppliance, using group measures of central tendency and trend, assessing the interaction of task and linguistic context. The chapter concludes with a discussion of the development of individuals, illustrating specific aspects of development over time and on different tasks. Chapter 6 provides a summary of the important findings of the study, as well as implications for future study and for second language teaching.

Chapter 2

Second Language Acquisition: Form-function Mappings, Variability and Development Over Time

Chapter 1 introduced the importance of the linguistic and psycholinguistic validity of instructional studies. To achieve such validity, it was argued that analyses must consider the development of interlanguage systems rather than just the target language. This chapter considers this issue in greater detail, considering functional interlanguage systems, the variability of second language use, and temporal changes in those systems and that variability. The chapter concludes with a set of six principles that integrate these issues and result in a series of 12 research hypotheses. The development and validation of these linguistic and cognitive principles is particularly important because such a theory provides a means of understanding and explaining the nature of instructed SLA and also provides the potential for developing further predictions regarding optimal types of instruction.

2.1 Alternative Approaches to Linguistic Analysis: Structural and Functional, General

Competencies and Interacting Sets of Form-function Mappings

Studies of language development may make very different assumptions about the nature of language and therefore may adopt very different types of language analysis. One important distinction is whether an analysis adopts a structural or functional approach to the study of language (cf. section 1.2). The structural and functional alternatives have often been considered and debated (e.g., Bates & MacWhinney, 1982, pp. 176-77; Andersen, 1990b). Although language has both structures and forms, and functions and meaning, linguists have often focused on either the formal or functional aspects of language. If an approach to SLA focuses on only one of these aspects of language, then the categories provided by that linguistic tradition will identify different research questions, use different methods and presumably find different answers and conclusions. Because language has both structures and meaning, these approaches can be

compatible. In fact, both views are necessary to obtain a complete understanding of the nature of language and language learning.

By considering both structural and functional linguistic categories, researchers may be able to better understand the effects of instruction on L2 development. If learners often do not achieve target-like use of articles or if instruction does not result in improvements, this may be due, for example, to the transfer of structural properties from the first language: many languages do not have articles or have article systems that are different in many respects from the English article system, and therefore a transfer of the L1 structural system can lead to non-target-like use. However, a second explanation of the difficulty of mastery may lie in examining the function that articles serve both in the target language and in the interlanguage system that the learner is developing. If the function of articles (i.e., the meaning they convey to a hearer/reader) is being conveyed by some other mechanism, then this may be the reason that learners are not improving. This second type of analysis is described here as either a functional analysis or an analysis of form-function mappings (cf. section 2.1.1) and was adopted in this dissertation.

An example of a structural analysis was reported in Mellow (1996), a preliminary study of the learners in this dissertation. Mellow (1996) adopted a structural approach and investigated whether the absence of articles in interlanguage could be attributed to the absence of functional categories in the interlanguage system (cf. Abney, 1987; Fukui & Speas, 1986; Grondin, 1992; Masterson, 1989; Ritter, 1992)². More specifically, the learners may not have had an interlanguage which required that each noun phrase (NP) be dominated by one but only one DP,

² A common assumption within principles and parameters syntactic theory is that there are two major classes of words and morphemes: (a) functional categories, including various grammatical morphemes, such as articles, possessive pronouns, the genitive suffix, auxiliaries, modals, certain verbal affixes, and complementizers (e.g., *that*, *for* and *if*); and (b) lexical or thematic categories, including nouns, verbs, adjectives and prepositions (e.g., Abney, 1987, pp. 54-68). Although certain languages, such as Japanese, may have forms that resemble functional categories, it has been argued that those forms may have properties that indicate that they are lexical categories. In addition, although the term *functional* is used for this class of language elements, principles and parameters theory is largely concerned with the formal or structural properties of language, and therefore the term *functional* has a very different meaning when used in this formal manner as opposed to the manner used throughout this study and introduced in chapter 1.

and that one and only D element must appear with each NP³. Instead, within the learners' interlanguage system, articles may have been one of a variety of optional pre-nominal modifiers, similar to adjectives. Although it is possible that all interlanguages begin without functional categories (as has been claimed for first language acquisition; Guilfoyle & Noonan, 1992; Radford, 1990), a less controversial proposal is that, when learning L2s with functional categories, learners would have difficulty if their first language, such as Japanese (Fukui & Speas, 1986), does not have functional categories. If it were true that these learners omitted articles because they had not acquired the distinction between functional and lexical elements, then this would have been reflected in the absence of other structural properties associated with functional categories, including iterability (permitting the presence of more than one D element; e.g., *John's the book, a your book*), and incorrect ordering (permitting elements such as adjectives to precede the D; e.g., *yellow the banana*). On the basis of the learners' responses to a judgement and correction task, Mellow (1996) argued that there was evidence that the learners had considerable knowledge regarding the properties of functional D elements. Therefore, the data in that study suggested that the omission of articles by these learners was not due to the absence of functional categories in their interlanguage.

Although the structural theory adopted in the analyses of Mellow (1996) allowed an assessment of specific properties of interlanguage (i.e., omission, iteration, and ordering of D elements) and provided a possible explanation of the difficulties that L2 learners have with English articles (i.e., the absence of functional categories in interlanguage), the results of that study suggested that, at least for these learners, there must be other causes of the difficulty with English articles. As a result, the study in this dissertation has adopted form-function analyses which provide insight into the omission of articles by L2 learners and provide a linguistically valid means of assessing the impact of instruction on language development.

³ Abney (1987, pp. 58-60) developed an influential syntactic analysis of determiners, suggesting that, in accord with the Projection Principle and X-bar theory, determiners and other similar morphemes form a category of functional elements, called D's, that *project to* (or are the head of) a syntactic structure known as the DP. DP and D therefore correspond roughly with the terms Determiner Phrase and Determiner.

A second distinction within approaches to linguistic analysis concerns the scope of the linguistic categories that are utilized. For example, many studies of instruction (e.g., many of those reviewed in Long, 1983) have used very general measures of language proficiency, such as scores from TOEFL tests and oral interviews. Other researchers (e.g., Harley et al., 1990; cf. Bachman, 1990, pp. 315 - 333) considered language proficiency to be composed of a small number of general competencies (e.g., grammatical, sociolinguistic, discourse, and strategic competencies).

Rather than one of these more general approaches to proficiency, this dissertation adopted the view, also expressed in Cumming and Mellow (1996) and Mellow and Cumming (1994), that language ability is composed of a large number of specific subsystems that interact in complex ways (cf. Huebner, 1979, p. 22). For example, knowledge of English use includes structural elements such as functional categories and projections (Abney, 1987; Fukui & Speas, 1986), language-specific articles and other pre-nominal elements (demonstratives, quantifiers, genitives, etc.), pronominalization, topicalization, and clefting. Knowledge of English use also includes functions such as the expression of old and new information, and the identification and tracking of referents. These and other subsystems interact as a set of form-function mappings and result in the specific properties of noun phrases (NPs) in English.

A valid interlanguage analysis therefore considers how any new linguistic knowledge has contributed to the functional system, for example, to the manner in which the learner expresses or packages intended meanings in order to communicate (e.g., Long & Sato, 1984, pp. 265-271; Tomlin, 1990; von Stutterheim & Klein, 1987). Such analyses may reveal that increases in the target-like use of one specific element may have been caused by a change that has altered a larger subsystem, resulting in non-target-like overgeneralizations about that element (cf. Huebner, 1979, 1983a, 1983b; Pica, 1983a), or that changes may not have occurred in the use of a specific element because the learner already utilizes another strategy, system, or structure to convey that meaning (cf. Long & Sato, 1984, pp. 271-273; Mellow & Fuller, 1995).

Thus, this approach suggests that an appropriate dependent variable for the study of the effects of instruction is a unit that is smaller than a general competency, such as grammatical competence, but is also larger than the use or suppliance of just one grammatical or discourse feature without reference to its role within a larger subsystem. Instead, an appropriate unit of analysis is a set of interacting form-function mappings. Determining the appropriate scope of the form-function mappings that are considered is a relative assessment that must be made and justified within the context of the purpose of each study. The remainder of this section supplies a justification for the scope of the form-function mappings considered in this dissertation study.

2.1.1 Functional Analyses of Language Use and Development

This section outlines an approach to language analysis that focuses on the functions of language elements rather than on their structures (e.g., Chafe, 1976, 1987; Givon, 1984). Functional approaches have also been described as conceptual, concept-oriented, meaning-oriented, communicative, and cognitive. Specific functional approaches to linguistics and language acquisition have been discussed by Bates and MacWhinney (1982, pp. 178-190), Klein (1990, p. 227), Klein and Perdue (1988, p. 3), Pfaff (1987, pp. 81-5), Tomlin (1990, pp. 158-62), and von Stutterheim and Klein (1987, pp. 191-3). The functional approach that is used in this study is a descriptive or relational approach (cf. Tomlin, 1990, pp. 159-60). In other words, this approach considers the possible *mapping relations* that hold between linguistic forms and semantic or pragmatic functions.

Two different relational or mapping approaches can be identified. In the first approach, researchers investigate the acquisition of functional systems such as topicalization or the tracking of referents. An example of this first approach can be seen in the work of Bates and MacWhinney (1982, pp. 174-175) who considered the first language (L1) acquisition of the functional system of topicalization or point making in discourse. Bates and MacWhinney examined the use and acquisition of various grammatical devices associated with topic and comment: a) topicalization devices such as assignment of sentence subject, pronominalization, ellipsis, and definite articles;

and b) commenting devices such as assignment of sentence predicate, specific lexicalization, indefinite articles, and connectors to previous discourse. Examining specific forms for tracking referents (i.e., distinguishing new, known, and current topics), Chaudron and Parker (1990) and Mellow and Fuller (1995) are also examples of the first type of analysis in SLA research.

The second type of functional approach to acquisition research seeks to understand how a learner acquires the correct function(s) of various morphemes or constructions (e.g., Klein 1990, p. 227). The second type of approach is therefore a functional analysis of structures, rather than an analysis of the structures that fulfill language functions. The second approach is very common in studies of the L1 and L2 development of article use (e.g., Maratsos, 1974; Huebner, 1983a, 1983b; Master, 1987; Thomas, 1989). This approach to examining form-function mappings is outlined below and is used in this study. For the purpose of this dissertation, the research questions are informed by the second functional approach: Within the context of the tracking of referents, what functions do articles serve in a learner's interlanguage and can this interlanguage system be affected by instruction? This is a particularly valuable approach because it examines development in a precise, explanatory fashion by considering how learners use these elements to convey meaning, rather than just examining language elements in reference to target-like structural properties.

2.1.2 Functional Analyses of Article Use and Development

Because of the considerable complexity of the functions of articles in English and of the complexity of the results of first and second language studies of article development, this section only provides an overview of several important trends in analysis and identifies several key issues which inform the instructional study in this dissertation. In English, articles convey a complex array of meanings. In addition to a number of idiosyncratic uses, differences in article choice (*the*, *a* or \emptyset) may reflect, for example: whether a noun is common or proper, mass or count, singular or plural, definite or indefinite, and generic or specific (cf. Celce-Murcia & Larsen-Freeman, 1983; Master, 1987). This complex set of meanings has been categorized in a variety of ways. For

example, some studies of article development, such as the *morpheme studies*, have considered this array of heterogeneous uses of articles as just one category (e.g., Bailey, Madden & Krashen, 1974; Brown, 1973; cf. Master, 1987, p. 7). As Huebner (1983b) and Pica (1983a), among others, pointed out, these morpheme studies tended to look at obligatory article usage (i.e., suppliance in obligatory contexts, SOC) and not at overgeneralizations in non-obligatory contexts, thereby excluding a considerable range of article usage (i.e., not fully assessing target-like use, TLU). Furthermore, as Master (1987, pp. 7 - 8) indicated, these early studies did not consider the zero article (\emptyset ; the required absence of a determiner in contexts such as before a proper noun, e.g., \emptyset *Steve broke the world record in \emptyset Oslo*). Excluding the zero article from analyses eliminates another significant portion of article usage.

Functional analyses do not consider article use as a single category, but instead sub-categorize article use in various ways. One simple distinction is between the definite article *the* and the indefinite article *a*, not distinguishing the sub-uses of each, and excluding the zero article which has an especially heterogeneous set of sub-uses (e.g., Cumming & Mellow, 1996). Other studies have distinguished between uses of *the*, *a* and the zero article (e.g., Yamada & Matsuura, 1982). In a revealing study to inform L2 pedagogy, Pica (1983b) analyzed article use in terms of a number of situational factors. Master (1987, pp. 6-25) summarized a number of other approaches to article classification that have been used in acquisition studies. An additional methodological issue is determining whether other pre-nominal elements should be included in an analysis. For example, Mellow and Fuller (1995) reported that, on a different task, the learners in this study used *this* and *that* where native speakers used *the*, and used *one* where native speakers used *a*, in accord with certain systematic patterns of use.⁴

Most of the recent functional L2 analyses of the acquisition articles have utilized a classification system first used by Huebner (1979; cited as a personal communication from Derek Bickerton; cf. Bickerton, 1981). In this system, noun phrase reference is divided into the four

⁴ In addition to the discussions of L2 development noted above, Ingram (1989), Maratsos (1979), Owens (1984), and Radford (1990, pp. 39, 44, 95-6) are examples of functional discussions of first language article development.

major categories in Table 2.1 below (cf. Master, 1987, p. 37). The following L2 studies have used some form of this classification: Huebner (1979, 1983a, 1983b), Huebner, Perdue and Carroll (1988, p. 36), Master (1987), Parrish (1987), Tarone and Parrish (1988), and Thomas (1989).

Table 2.1

Bickerton's Four Categories of Article Use, from Huebner (1979), with Examples Based upon Tarone and Parrish (1988, p. 27)

1. generics (-specific referent, +assumed known to the hearer); articles: *a*, *the*, \emptyset ;
e.g., The lion is a beautiful animal.
2. specific definite (+specific referent, +assumed known to the hearer); article: *the*;
e.g., He married a woman from Scotland. The woman is from Gress.
3. specific indefinite (+specific referent, -assumed known to the hearer); articles: *a*, \emptyset ;
e.g., She married a man from the Isle of Harris.
4. indefinite generic (-specific referent, -assumed known to the hearer); articles: *a*, \emptyset .
e.g., I don't see a pencil.

Several studies have expanded on this classification system, with Master (1987, p. 40) using 19 categories and Thomas (1989) using 21 categories. These categories include classifications such as: generic *a*, generic *the*, generic \emptyset , subsequent mention *the*, shared knowledge *the*, count singular *a*, noncount \emptyset , interrogative *a*, etc. (Master, 1987, p. 40). Reference grammars, such as Celce-Murcia and Larsen-Freeman (1983), typically provide a description of these functional categories. Other functional analyses of article development are reported in Chaudron and Parker (1990), Fakhri (1989), Givon (1984), Kumpf (1992), Mellow and Fuller (1995), and Pfaff (1987a).

These studies of the SLA of articles have suggested a number of interesting developmental trends. In early stages of development, articles are rarely used, especially for learners whose first language lacks articles (e.g., Chaudron & Parker, 1990, p. 57; Huebner, Perdue & Carroll, 1988, pp. 45-6; Master, 1987; Thomas, 1989, p. 340). Perhaps the most fundamental developmental trend that occurs after the early omission of articles is that as learners gradually become more proficient with the L2, the use of articles becomes more target-like. In fact, both Master (1987) and Cumming and Mellow (1996) have reported evidence which indicates that article development

in general may be one of the most valid indicators of the development of overall proficiency in English.

However, the functional analyses in these developmental studies have indicated that the increases in target-like (TL) use of English articles do not proceed in a simple linear fashion, with the correct use of *the* and *a* gradually replacing the omission of those articles (i.e., the overgeneralization of \emptyset). Instead, development appears to be characterized by a number of interesting stages and influences. Table 2.2 below indicates a hypothetical or idealized developmental sequence for the use of articles with specific reference noun phrases. It is important to emphasize that these stages are hypothetical, based upon previous developmental data: a learner will not necessarily go through all of the stages, nor will the stages necessarily be clear and distinct for any given learner.⁵ Instead, the stages in Table 2.2 are provided to indicate the complexity of the developmental data.

Table 2.2
Hypothetical Development Stages in the Acquisition of Article Use with Referential Noun Phrases

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5 (TL)
definite NPs	\emptyset	the	the	a	the
indefinite NPs	\emptyset	\emptyset	the	a	a / \emptyset

As indicated in Table 2.2, the first stage of development is often one in which all articles are omitted. The next two stages indicate the general trend in which referential definites develop before indefinites (or in which elements to indicate known or old referents emerge before those that indicate new referents; cf. the discussion of markedness in Chaudron & Parker, 1990, pp. 47-49). Stage 2 indicates that *the* is often used for definite noun phrases and that indefinite noun phrases

⁵ In addition, this chart does not indicate all of the possible developmental trends. For example, Huebner, Perdue and Carroll (1988, pp. 45-46) indicate that at an early stage before *the* develops, *one+noun* or *quantifier+noun* may be used systematically to disambiguate noun phrases (cf. Mellow & Fuller, 1995).

are marked by \emptyset (e.g., Chaudron & Parker, 1990, pp. 57, 59; Huebner, Perdue & Carroll, 1988, p. 58; Tarone, 1988, p. 106). This usage is grammatical for indefinite mass and plural nouns and might be characterized as an overgeneralization of \emptyset use to singular indefinite nouns. Stage 3 is often described as *the-flooding*, the overgeneralization of *the* to all specific reference noun phrases (e.g., Chaudron & Parker, 1990, p. 59; Huebner, Perdue & Carroll, 1988, p. 101; Master, 1987; Thomas, 1989, esp. 340).⁶ Although the flooding of *the* may retreat directly to a target-like stage 5, some learners might flood or overgeneralize the use of *a* when it first emerges in their interlanguage, as represented by stage 4 (e.g., Chaudron and Parker, 1990, p. 59).

These three potential intermediate stages (2, 3, and 4) illustrate the importance of functional analyses that involve several sub-categories of article use and examine the interaction of these specific elements within a larger subsystem. In other words, a change from one stage to another indicates development in the interlanguage system that may not be reflected in an increase in the percentage of target-like use for all articles combined in a single category. Although the analyses in this study do not consider these stages or the relationships between article use in Bickerton's four-category system, it may be that developments such as these can account for some of the variation that is not explained by the categories of analysis that are used.

2.1.3 Functional Analyses of the Suppliance for Tracking Specific Referent Noun Phrases

For the purposes of this investigation, the scope of inquiry was restricted to a particular subset of English article use. The analyses in this dissertation were only concerned with the articles that are used with common nouns, including mass nouns and singular and plural count nouns, that refer to specific referent (+SR) noun phrases that are assumed known to the hearer (+HK) (i.e., category 2 in Table 2.1). In English, *the* is the only article that is used for these *referentially definite* noun phrases.

⁶ This overgeneralization may extend to generic noun phrases as well (e.g., Huebner, 1983; cf. Thomas, 1989).

The choice of this subset of article use permitted an analysis that was not too ambiguous, complex, or unreliable, and required only a small unit of instruction. This analysis excluded proper nouns, idiosyncratic uses of articles, and noun phrases with referents that are not assumed known to the hearer or are generic. This exclusive focus omitted, as noted above, many important aspects of article use and development. At the same time, the area was broad enough to include several meaningful and relevant aspects of language use. This choice also made it possible to create parallel versions of the measurement instruments and permitted greater reliability in the transcription and coding of the data, issues that are discussed in Chapters 3 and 4 and are very important for permitting the assessment of development.

The use of articles with specific referent noun phrases is of particular interest because these form-function mappings are one of several linguistic means of tracking entities or referents in discourse. Referent tracking is the use of a system of linguistic mechanisms to indicate whether a referent (the entity expressed by a noun phrase) is new to the discussion, was mentioned much earlier in the discussion, was mentioned recently but needs to be distinguished from other entities (especially semantically similar entities, such as those that are equal in terms of animacy, potential for agenthood, number, gender, etc.), or was mentioned recently and does not need to be distinguished from other entities⁷. Several of these distinctions are illustrated in 6).

- 6) A woman is standing beside a girl. The woman is talking to the girl. The girl is nodding her head.

The first mention of each referent (*woman* and *girl*) is preceded by the article *a*. In the subsequent mention of those referents, they are preceded by the article *the*. In the final sentence, a full noun phrase with an article is used (*The girl*) instead of a pronoun (*she*), presumably to clarify the referent and indicating who was nodding her head.

⁷ There are other possibilities for tracking referents as well, for example, the referent is logically inferable from something earlier in the text or is evoked by something in the situation in which the speech event is taking place, that is, the physical presence of the entity, perhaps in a picture (cf. Prince, 1981).

The tracking of NP reference overlaps with a number of other means of categorizing noun phrases, many of which discuss the status of information, and which have been described using concepts such as predictability, expectedness, accessibility, and topic continuity (e.g., Chafe, 1976, 1987; Chaudron & Parker, 1990; Givon, 1983; Klein & Perdue, 1988; Mellow & Fuller, 1995; Prince, 1981; Yule, 1981). The following sets of terms (a non-exhaustive list) are used for similar although not identical purposes: definite / indefinite; old / new; given/ new; current / known / new; current / displaced / new; assumed known / assumed unknown; already active / previously semi-active / previously inactive; accessible / not accessible; continuous / non-continuous; predictable / not predictable; expected / unexpected; and textually evoked / situationally evoked / un-used / brand new. In this dissertation, the term *predictability* (rather than *topic continuity*) is used to describe the status of a NP with regard to the tracking of referents. Predictability includes both referential distance (e.g., the number of clauses since the last mention of the referent) and ambiguity due to competing referents (cf. Givon, 1983, pp. 111-112). In this dissertation, the primary categories that are used to distinguish the predictability of noun phrases are: *assumed known* and *new*. In addition, to distinguish degrees of referential distance, the following categories are used: *current*, *recent*, and *old* (all of which are assumed known). Current referents are those that were mentioned in the previous clause. Recent referents are those that were last mentioned two to five clauses earlier in the discourse. Old referents are those that were last mentioned more than five clauses earlier in the discourse. These categories are discussed in greater detail in Chapter 4.⁸

As illustrated in 6) above, articles are one of many devices that may be used to track referents. For example, if a referent is new to the discussion, has not been discussed in some time, or needs to be distinguished from another entity, the speaker may use several devices to draw

⁸ To clarify the focus of the discussion, predictability within the discourse is being considered, leading to the continuum in (a), rather than a scale of time of introduction to the text, as indicated by the continuum in (b):

- (a) least predictable - new/old/recent/current - most predictable
- (b) newest introduction - new/current/recent/old - oldest introduction

attention to the fact that the referent is relatively unpredictable. The English mechanisms that indicate unpredictability include the use of the full noun phrase rather than a pronoun, and several word order variations (existentials, for new referents; left-dislocation, for old referents, mainly used in colloquial and conversational English, Givon, 1983), illustrated in 7) and 8) below.

- 7) existentials - *There is a woman.* . . .
 8) left-dislocation - . . . *The woman*, she is talking to the girl.

In contrast, if a referent has recently been discussed or does not need to be distinguished from other referents, then other devices can be used. The English mechanisms that indicate *predictability* (or continuity) include pronouns and zero anaphora, illustrated in 9) and 10) below.

- 9) pronouns - There is a woman. *She* is talking to the girl.
 10) zero anaphora - The woman is talking to the girl and \emptyset holding a book.

These examples indicate that articles are generally used only with relatively unpredictable NPs. These distinctions are roughly summarized in 11).

- 11) unpredictable, new referents; never mentioned in the text
 mechanism: indefinite article and noun
 example: *a woman*
- unpredictable, old referents; not recently mentioned in the text
 mechanism: definite article and noun
 example: *the woman*
- predictable, current referents; very recently mentioned
 mechanisms: pronoun or zero anaphora (definite article and noun
 are possible for certain purposes)
 examples: *she*, \emptyset

The summary in 11) indicates that pronouns or zero anaphora, rather than full noun phrases with articles, are often used for more predictable NPs. One important implication of this discussion is that articles are only one of several means of tracking nominal reference and therefore articles may not be used by L2 learners if they are relying on other mechanisms to express certain meanings (e.g., Mellow & Fuller, 1995).

An examination of the linguistic system for tracking referents reveals that, in addition to the developmental patterns described in section 2.1.2, a number of other factors affect the L2 use of

articles. These factors may explain, in a systematic fashion, why learners sometimes omit the article *the* and use only a bare noun for referentially definite (+SR, +HK) NPs. Two of these factors involve an examination of the interaction of discursual and cognitive processes involved in form-function mappings: the use of *the* with respect to the position of the noun phrase within the sentence (a formal aspect of language use) and the use of *the* with respect to referential distance (a functional aspect of language use).

2.1.3.1 *the Use with Respect to the Position of the Noun Phrase Within the Sentence*

The preceding discussion of referent tracking indicated that English has a number of forms or structures for indicating the predictability of a referent, for example, both existential constructions and the use of a full NP with *a* indicate a new referent. In addition, it appears as though some or all second language learners may have an interlanguage system that includes an additional structural means of expressing predictability, that is, a *topic-comment* system. In contrast to the subject-predicate system of English, a topic-comment system is a word order pattern in which the first part of an utterance (often a noun phrase) is the topic, usually already known to the hearer, and the second part of the utterance is a comment about that topic, potentially providing new information about the topic (for discussions of differing definitions of topics and related terms such as theme and rheme, see, for example, Brown & Yule, 1983a; Duff, 1985; Finegan, 1994, pp. 200-201; Huebner, 1983b; McCarthy, 1991, pp. 51-59, 131-136). Although clear examples of utterances that indicate a topic-comment system might resemble the left-dislocated structure in 8) above (cf. Larsen-Freeman & Long, 1991, p. 267), interlanguage utterances of the form [NP₁ Verb NP₂], which appear to have English [Subject - Verb - Object] structure, might also have topic-comment properties. In this dissertation, the target-oriented terms, *subject* and *non-subject* will be used to describe the first NP position and subsequent post-verbal NP positions, respectively, in the interlanguage data. The terms *subject* and *non-subject* are used for reasons of simplicity, because the data is largely of the form [NP₁ Verb NP₂] and therefore the term *topic position* is used only for left-dislocated English interlanguage structures. This choice of terms is

not intended to minimize the importance of a functional interlanguage perspective and does not imply the absence of topic-comment properties for those NPs.

For two reasons, it is hypothesized that Japanese ESL learners may have a topic-comment system in their interlanguage. First, Japanese has topic-comment structures (Duff; 1985, pp. 33-35; Shibatani, 1990, p. 274) and therefore these learners may transfer this system to their interlanguage (cf. Huebner, 1979, pp. 26-7). Second, it may generally be true that interlanguages, especially in their early stages, have topic-comment systems (e.g., Chaudron & Parker, 1990, p. 50; Duff, 1985; Givon, 1983; Huebner 1983a, 1983b; cf. the review by Larsen-Freeman & Long, 1991, pp. 74-75, 267-268). In addition, even if these learners do not have dominant topic-comment systems, and instead have systems that are more syntactic, with subject-predicate relations, the first NP in their sentences may nonetheless be assumed known to the hearer because in English the first NP (the subject) is usually assumed known to the hearer, while new referents appear most often as objects rather than as subjects (e.g., Chafe, 1987, p. 37; Prince, 1992).

If learners have topic-comment structures (or properties) in their interlanguage, then this suggests that placing a referent in the subject (or first noun phrase) position in a sentence or utterance is an additional linguistic means of indicating that the referent is assumed known to the hearer. Consequently, the use of *the* is actually redundant for assumed known noun phrases in subject position in interlanguage.⁹ The importance of this multiple marking of the assumed known [+HK] status is discussed further in section 2.4, where the differential suppliance of *the* in subject and non-subject positions will be discussed. In addition, although not investigated in this study, new referents may be redundant when they appear in non-subject or comment positions.

2.1.3.2 the Use with Respect to Referential Distance

Whereas section 2.1.3.1 considered an additional form that conveyed the same function as the presence of *the*, this section considers an additional function that is conveyed by *the*. An

⁹ In structures that are more clearly topic-comment, with left-dislocation as in 8), the use of *the* is also redundant.

examination of the English forms used in tracking referents, briefly summarized above, indicates that speakers tend to use more linguistic material (e.g., full NPs, existentials, left dislocation) to mark relatively less predictable referents, and less material (e.g., pronouns, zero anaphora) to mark predictable referents. Givon (1983, p. 126) has referred to this as the Quantity Universal: “More continuous, predictable, non-disruptive topics will be marked by *less marking material*; while less continuous, unpredictable/surprising, or disruptive topics will be marked by *more marking material* (original italics retained).”

As noted above, *the* suppliance before a noun can be used for any assumed known NP, contrasting with the use of *a* for new referents. However, an interesting possibility is that *the* will be supplied differently in response to an important aspect of predictability, that is, referential distance (e.g., number of clauses since the last mention of the referent) (cf. Chaudron & Parker, 1990; Huebner, Perdue & Carroll, 1988, p. 59; Mellow & Fuller, 1995). In accord with Givon’s Quantity Universal, one might expect that suppliance of *the*, which is more marking material, is used for more referentially distant NPs (i.e., old and recent NPs), and that less material, the omission of *the* resulting in a bare noun, is used for NPs that are less referentially distant (i.e., current NPs). Thus, for current referents, suppliance of *the* may fulfill two functions: to contrast with referents assumed unknown to the hearer (for which *a* is used), and to express referential distance.

The possibility that the Quantity Universal affects the tracking of referents in interlanguage has been investigated by Chaudron and Parker (1990, p. 57). Their data suggested that, in comparison to full NPs with definite articles, bare nouns were used more often for current rather than known topics, and that bare nouns are used more often at lower proficiency levels. Chaudron and Parker’s results and the implications of the polyfunctionality of *the* are explored in section 2.4, providing a research hypothesis for this study.

2.2 Variability in Interlanguage: Task Variation and Individual Variation

In order to inform a valid study of the effects of instruction on interlanguage development, the previous section (2.1) considered the development of form-function mappings, revealing valuable categories of analysis. Language variation is a second important issue that affects language use, and especially appears to affect interlanguage use. This section will first consider task variation and then will consider the nature of individual variation.

The nature and importance of language variation is interpreted quite differently by different approaches to linguistic theory. For some theories, variation is peripheral to central linguistic issues because language competence is thought to be homogeneous. For example, following Chomsky (e.g., 1965), some SLA researchers (e.g., Gregg, 1990; cf. White, 1989) assume that learners have an abstract and uniform knowledge of language (competence) and that any variation in language use is due to performance problems, for example, slips of the tongue, distractions, boredom, fatigue, etc. (see Radford, 1981, pp. 2-4 for further discussion). A homogeneous view of language ability implies that all types of language tasks might reflect linguistic knowledge, and therefore metalinguistic measurement instruments, such as grammaticality judgement and correction tasks, are appropriate tools of investigation. Mellow (1996) used this type of measurement instrument to assess the competence of the same learners in the present thesis.

In contrast, a heterogeneous view of language competence (cf. Labov, 1970) has been adopted by other SLA researchers. These researchers argue that interlanguage is systematically variable and are interested in the ways in which language is used differently in different contexts (e.g., Ellis, 1985, 1989; Preston, 1989; Tarone, 1988; Young, 1989). The heterogeneous view is often adopted by pedagogical researchers: Many language educators are interested in how learners actually use language in various contexts, and are less concerned with the abstract rules and knowledge that learners possess. Thus, although metalinguistic performance is one aspect of language use, other concerns may be especially relevant and valid for pedagogy. The study in this dissertation adopts the heterogeneous view of language, assuming that language use varies with

respect to the context in which it is used. As a result, this study investigates whether instruction may have different effects when language use is assessed on a variety of tasks.

A number of linguists and applied linguists have provided extensive lists or frameworks which attempt to summarize the factors affecting the variable nature of language use (e.g., Bachman, 1990, pp. 111-157, 163-165, esp. 119; Ellis, 1985, pp. 119-120, 1989; Labov, 1970; Preston, 1989, esp. 218-219; Tarone, 1988; Wells, 1986, pp. 112, 117-133; Young, 1989, pp. 72, 76-79). These factors are reflected in and can be used to classify the many different types of tasks that have been used to measure language, including grammaticality judgements (potentially with measures of response time); correction of words, sentences, and texts; multiple choice; fill in the blanks (for single sentences); cloze tests (for longer texts); imitation; story retelling; story telling; picture description; directed questions; essay writing; spontaneous interviews; hidden microphones during play or talk; among many others. Because the factors which affect language use are vast and not yet well understood (Wells, 1986, p. 114), rather than attempting to summarize or synthesize the large number of factors listed in the above mentioned references, I will discuss a specific subset of those factors that will be investigated in this study.

Given the aspects of interlanguage use that are investigated in this study, the *amount of attention paid to speech* and its effects on language variation are of considerable interest. Labov (1970) hypothesized that the amount of attention paid to speech is the crucial factor causing variability: more formal styles of speech require more attention to speech. This hypothesis dominated early SLA research on variability (Ellis & Roberts, 1987, pp. 14-15). However, attention paid to speech as a variable has been criticized by a number of researchers (e.g., Crookes, 1988, pp. 21-23; Eisenstein & Starbuck, 1989; Labov, 1970, p. 70; Lantolf & Ahmed, 1989; Rampton, 1987, pp. 48, 61; Skehan, 1987, pp. 20-23; Swan, 1987, pp. 61-62; Tarone, 1985, p. 375fn, 1988, pp. 42-45). The main criticisms are that (a) it is very difficult to quantify or control the amount of attention paid to aspects of language, and (b) attention to form is not an explanatory factor: Researchers still need to determine what it is in a task that causes this attention to form (e.g., assumed expectations and impositions; emotional investment; complexity and familiarity of

content, grammatical items, discourse properties, rhetorical structure, and vocabulary). In addition, attention to aspects of language may occur at two times: prior to speech (planning) and in controlling the execution of speech (monitoring) (Crookes, 1988, 1989). As a result, if the factor of attention to speech is to be used for systematic inquiry, it needs to be made more precise and some of the factors which affect attention need to be stipulated and controlled.

Attention to speech, as an important factor affecting variability of language use, is considered in this dissertation for three reasons. First, following the Labovian tradition, especially as developed in SLA research by Tarone (1985, 1988) and Tarone and Parrish (1988), one of the hypotheses investigated in the dissertation is that forms will be supplied in the most target-like manner on tasks with the most attention to form, and in the least target-like manner on tasks with the least attention to form (see Hypothesis 1 in section 2.5).

The second reason for considering attention to speech in this instructional study is to address concerns of pedagogical validity. As indicated in section 1.3, one of the concerns of educators is to assess the effects of instruction on spontaneous language use outside of the classroom context. However, because the tasks used to measure language ability are usually somewhat contrived and suffer from the observer's paradox (Labov, 1970, p. 47: observing the learner changes the learner's behavior), the tasks chosen for a research study may not necessarily represent authentic language use. For example, in this and many instructional studies, the students knew they were part of an experiment and presumably behaved differently as a result. Therefore, the decision in this dissertation was to design tasks that could assess changes in language performance in contexts in which the learner is not likely to rely upon conscious, analysed, non-automatic, declarative, metalinguistic form-focused knowledge. Thus, this study is designed to assess whether instruction results in the automatization of linguistic knowledge (e.g., McLaughlin, 1987, 1990; cf. Krashen, 1992; see section 1.3). As a result, one particularly important aspect of investigating variability for the present study is the use of measurement instruments or tasks in which the learners are relatively unlikely to modify their use of language as a result of attention to

specific forms and the use of conscious rules about those forms (cf. Ellis, 1989, pp. 33-4, 40-1; Hulstijn & Hulstijn, 1984; Tarone, 1988, pp. 33-45) (see Hypotheses 6, 7a and 7b in section 2.5).

The third reason for considering attention to speech in this instructional study is to investigate theoretical predictions regarding the effects of the availability of attentional resources during the performance of a task. Specifically, Mellow and Cumming (1994) have indicated that certain cognitive processes, such as the priming of redundant elements, occur during tasks such as writing, in which learners have relatively abundant attentional resources available for focusing on grammatical forms, but that other cognitive processes, such as the efficient deletion of redundant elements, occur during tasks such as speaking while attending to content, in which learners have relatively limited attentional resources (cf. VanPatten, 1993, 1994). Mellow and Cumming have argued that task differences in the availability of attentional resources result in specific types of asymmetries in interlanguage use. These asymmetries will be explored further in section 2.4 (see Hypotheses 2, 3, 8a, 8b, and 9 in section 2.5).

As a result, the present study includes a range of tasks in order to determine if both conscious knowledge and more automatic language behavior change as a result of instruction. In particular, one task in this study attempted to elicit language in a context in which the focus is primarily upon meaning and in which little time is allowed to draw upon any conscious grammar rules (cf. the discussion of the oral fable retelling in section 4.1.4). Because of the criticisms of research on attention to speech, a detailed discussion of the efforts taken in this study to control attention to form is provided in sections 4.1.3 and 4.1.4.

In addition to the issues of variation discussed above, which considered task variation as a result of attention to speech, an additional aspect of variation is the nature of individual variation. The behavior of individuals might be studied for two different purposes: (a) to assess whether the patterns found in group behavior can be generalized to the behavior of individuals; and (b) to investigate variations in the developmental course of instructed SLA. The first issue, that results found in groups may not be generalizable to specific individuals, is an aspect of external validity that has been discussed by other researchers (cf. Kratochwill, 1978, pp. 23-24; Mellow & Fuller,

1995; Mellow, Reeder, & Forster, in press; Runkel, 1990, esp. pp. 3-5, 167-168). This issue is particularly important in the study of SLA because there appears to be considerable individual variation in how languages are learned, for example, variations in first language, learning style, motivation, attitudes toward the target culture, educational background, etc. (e.g., Ellis, 1994; Larsen-Freeman & Long, 1991). Thus, instruction that is beneficial, on average, for a group may not actually benefit all or even any of the specific learners with whom teachers are working. As a result, one avenue of investigation in an instructional study is to consider whether the typical behavior of a group is an accurate indication of the behavior of each of the individuals in the group. A preliminary investigation of the complex relationship between group and individual behavior is provided in section 5.3 (see Hypothesis 12 in section 2.5). Mellow and Fuller (1995) discuss the issue of group and individual behavior in further detail.

The second purpose noted above, to investigate individual variations in the developmental course of instructed SLA, is the primary focus of section 5.3. Thus, individual behavior is considered in detail, in reference to theories of SLA, in order to illustrate hypothesized patterns of instructed SLA. These patterns may involve temporal variation, manifestations of how instruction affects the language use of individuals over time (cf. section 1.4). Specifically, temporary and lasting effects, as well as immediate, delayed, gradual, and U-shaped effects will be considered (see Hypothesis 12 in section 2.5). These issues are discussed in greater detail in the next section.

2.3 A Theory of the Developmental Course of SLA

The previous two sections have considered the nature of form-function mappings and variability. This section considers a third important aspect of interlanguage development, the nature of changes over time, especially in terms of possible responses to instruction. Mellow, Reeder, and Forster (in press) indicated that educators and theorists are most concerned with changes in spontaneous or automatic language use, and permanent or lasting changes (cf. Handscombe, 1990, pp. 183-184; Krashen, 1993; Lightbown & Spada, 1990, pp. 432, 444 fn., 1993, p. 102; Mellow, 1992; Spada & Lightbown, 1993, p. 208). In other words, educators

wish to know whether specific sessions of instruction cause interlanguage knowledge to become fairly permanently automatized or restructured. To do this, changes in the language use of learners must be measured on tasks that permit varying degrees of attention to form (i.e., monitored and unmonitored tasks) and many data samples must be longitudinally collected in order to understand how language use may change over time. Such data may validly measure a construct that can be referred to as the *developmental course of instructed SLA* (cf. Mellow, Reeder & Forster, in press).

In this dissertation, the term developmental course is used to refer to language development over time, including both acquisition and attrition, and on different tasks. The term course is being used, rather than terms such as *route*, *stages*, *sequence*, or *order*, because these other terms have fairly established meanings within the SLA literature. The construct of developmental course may be the central component of a transition theory of SLA (cf. Long, 1993; Gregg, 1989, 1993). As Gregg (1989, p. 31) suggests, SLA theory needs more than "a proliferation of terminology" and "a lot of flow charts and diagrams". To assess this construct, research methods must be able to evaluate temporal changes, including changes that may be immediate, gradual, delayed, incubated, or residual (e.g., Cohen, 1975, p. 137; Ellis, 1990a, p. 168-70; Gardner, Lalonde & MacPherson, 1985; Hedgcock, 1991; Larsen-Freeman, 1991, p. 336; Mellow, 1992; van Els, 1986, p. 10; Weltens, 1987, pp. 29-30). Processes such as restructuring or automatization may be involved in this transitional process (cf. McLaughlin, 1990). In addition to assessing the onset of instructional effects, research also needs to evaluate the permanency of those effects, whether they last for hours, days, weeks, months, or years. Because attrition or forgetting is an elemental and often inevitable constituent of the learning process (e.g., Andersen, 1982, pp. 84-6; Hedgcock, 1991, 1994; van Els 1986, pp. 8-9; Weltens & Cohen, 1989), an understanding of the effects of instruction requires an understanding of the conditions under which loss may occur and the rate at which loss may occur. Furthermore, as suggested in the previous section, research methods need to assess variable changes in language use, including the possibility of different developmental patterns on different types of tasks. Thus, the development of ability on monitored,

focus-on-form tasks, such as grammaticality judgements, sentence corrections, or cloze tests, may be different from the development of ability on tasks that require less monitored language use, such as oral interviews, story retellings, or picture descriptions (cf. Ellis, 1989; Tarone, 1988, 1993). Overall, investigations of these temporal and variational concerns will operationalize and verify the pedagogically and theoretically relevant construct of the developmental course of instructed SLA.

2.3.1 The Temporal Nature of the Developmental Course of Instructed SLA

In order to better understand the effects of instruction on interlanguage development, the temporal nature of this developmental course must be considered. This section will examine possible temporal learning curves, and their implications for instructional studies of lasting effects. A variety of possible learning curves, in response to instruction, are illustrated in Figure 2.1. In each of the ten graphs, time is represented by the horizontal *x*-axis and developments in language use are represented by increases upward along the vertical *y*-axis. For the purposes of illustration and because many instructional studies use a measure of accuracy, the *y*-axis in Figure 1 is labeled "Accuracy". However, as noted in previous discussions of the development of form-function mappings, the variable, nonlinear nature of interlanguage development implies that overall accuracy may not be as valid a measure as is a consideration of the use of form-function mappings in a range of different contexts (cf. Larsen-Freeman & Long, 1991, pp. 40-41; Pica, 1983a). The thick vertical line in each graph in Figure 2.1 represents a session or phase of instruction. These idealized graphs have relatively smooth curves and do not illustrate the variable nature of language use, that is, fluctuations in accuracy. In graphs 6 through 10, the thin line in the postintervention phase represents the expected development of the learners receiving treatment, as would be indicated by the preintervention baseline trend (or by a control group, in the case of postintervention changes in trend, such as in 10). In all graphs except 5 and 8, a clear developmental trend in the preintervention baseline has been included. The first three graphs represent cases in which instruction appears to have a detrimental effect on language use. In

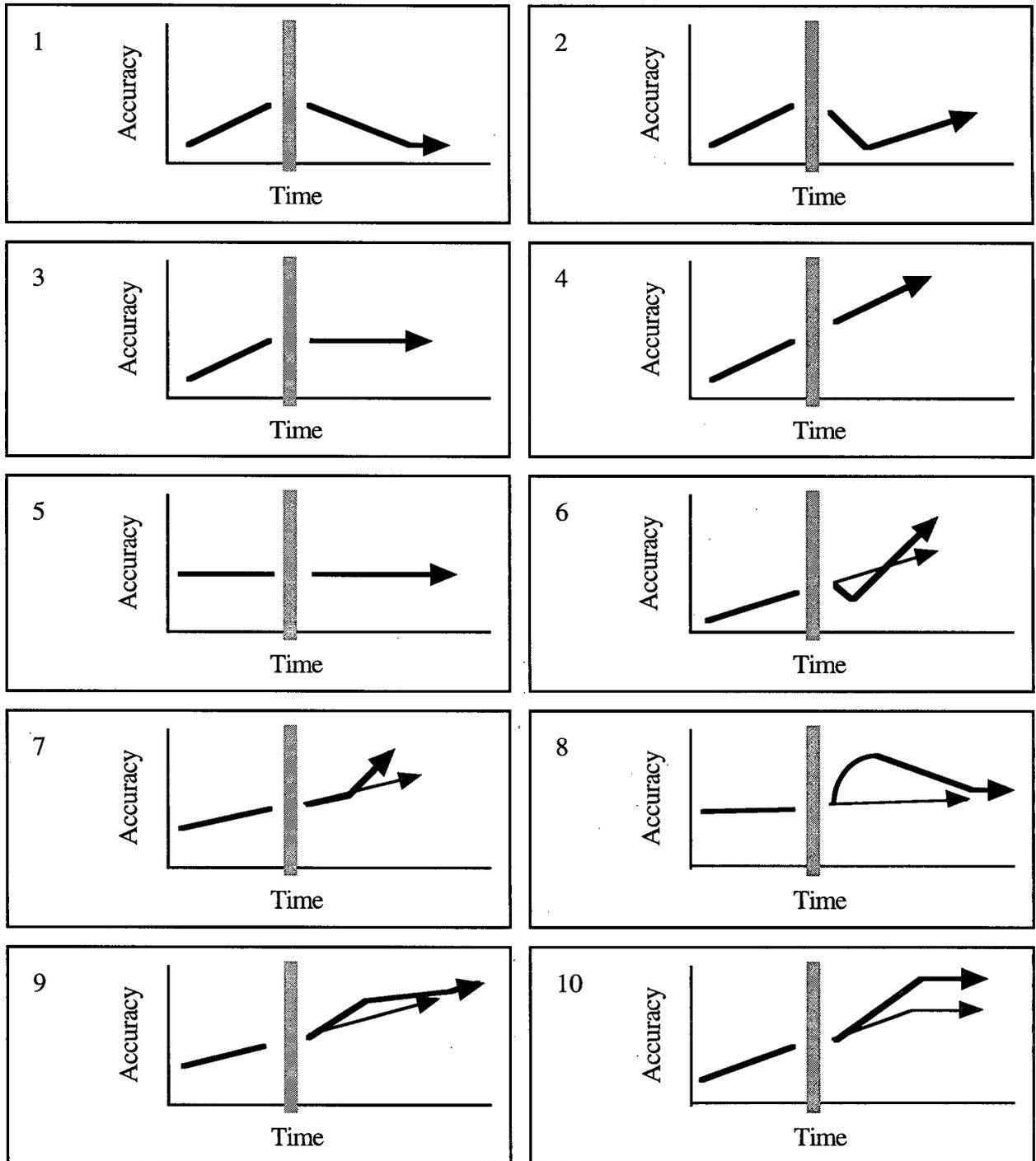


Figure 2.1 Possible temporal learning curves

graphs 4 and 5, instruction has no effect on the developmental course. Graph 6 represents a case in which instruction leads to restructuring followed by significant improvement.

Graph 7 represents a case in which improvement occurs after a delay. This possibility is supported by the findings of attrition research (e.g., Cohen, 1975). As summarized by Oxford (1982, pp. 167-168) and Weltens (1987, pp. 29-30), this research indicates that residual learning may continue for at least 3 months after instruction, although not for all language items, nor for all individuals.

Graphs 8, 9 and 10 provide insight into the potential for lasting effects of instruction. In graph 8, instruction results in immediate improvement, but those gains are then lost and performance returns to the preintervention baseline level. This type of result, temporary improvement, was obtained in White (1991), with posttests given one year after instruction. Because attrition is a common aspect of the developmental course, researchers might expect to find this type of learning curve, especially (a) for foreign language learners who do not have the benefit of practicing the language outside of the classroom (making permanent effects extremely unlikely); (b) for infrequent elements, such as adverbs or marked relative clauses, which the learner may rarely use after the instructional session; and (c) when much time has elapsed since instruction, perhaps 6 months or more. Beyond 6 to 8 months, attrition studies demonstrate the loss of language use, although this is dependent on the original proficiency of the learner, among other factors.

In graph 9, instruction results in immediate improvement, but then the rate of improvement decreases, and the baseline predictions (or control group), with steady improvement, eventually achieve the same performance as the instructed learners. Like attrition, decreases in the rate of improvement are expected, in this case because a ceiling effect results when the learner approaches a maximal proportion of a target-like behaviour.¹⁰ This type of result, an increase in rate of

¹⁰ The ceiling may not necessarily be at 100% of the target-like use of a dependent variable, since that variable may actually assess a language ability that is affected by more than one factor. The need for precise dependent variables was discussed in section 2.1 with regard to the scope of linguistic categories.

development with no lasting effects, was obtained in Harley (1989a), with post-tests given 13 weeks after instruction.

In graph 10, instruction results in immediate improvement and even though the rate of improvement then decreases, as it must, the control group also stops improving, stabilizing at a level below that achieved by the instructed learners. This type of result, an increase in rate of improvement and in level of ultimate achievement (at least for mid-term effects), was obtained in Day and Shapson (1991), with posttests given 11 weeks after instruction. A comparison of graphs 8, 9, and 10 reveals that lasting effects correspond to an increase in the ultimate level of attainment and that temporary (short- or mid-term) effects correspond to an improvement in the rate of development. Importantly, this comparison also indicates the only conditions under which long-term effects will be obtained: when a control group stabilizes at a relatively low level *and* when attrition does not occur in the instructed learners.

This examination of developmental patterns has a number of implications for the present study of the effects of instruction, especially for assessing lasting effects (for further discussion, see Mellow, Reeder & Forster, in press). First, a control group design is required to establish any long-term effects of instruction, because it must be established that the learners would otherwise have stabilized (although establishing the equivalency of that control group may be difficult). For this reason, a control group has been used in this study, although the challenges involved in using non-equivalent groups require caution in the interpretation of results. Second, because a control group may not stabilize, long-term or permanent effects may not occur, especially for the types of learners in this study: those who are studying in classrooms and who are living where the target language is spoken (cf. the discussion of developmental trend in Mellow, Reeder & Forster, in press). As will be seen in sections 5.2 and 5.3, the continuing development by the control group, or by a subset of the control group, resulted in effects that were not statistically significant.

Although several of the developmental courses proposed in Figure 2.1 have been attested, the details of those and other learning curves need to be empirically verified. Therefore, a third implication for this study is that the actual time frame and shape of those learning curves requires

investigation. The nature of potential changes that can be investigated include the rate of initial improvement, the possibility of delayed learning, and the length of the time until attrition may occur. In addition, complex aspects of learning can also be investigated, including possibilities of delayed or cyclic development (e.g., Cook & Campbell, 1979, pp. 227-28; Tawney & Gast, 1984, pp. 71, 118-9), automatization and restructuring (e.g., McLaughlin, 1990), and U-shaped behavioral growth (e.g., Reeder & Shapiro, 1993; Strauss, 1982). Other complex issues that could potentially be addressed include (a) distinguishing between attrition and a failure to acquire (Andersen, 1982, p. 85; van Els, 1986, pp. 8-9); (b) assessing the similarities and differences in developmental and attritional sequences (e.g., Hedgcock, 1991; Kaufman, 1992); (c) assessing the impact of the many factors that affect both attrition and instruction (e.g., Ellis, 1990a; Andersen, 1982; Oxford, 1982; Lambert & Moore, 1986; Weltens, 1987); (d) understanding why attrition does not always occur (e.g., de Bot & Weltens, 1992); and (e) assessing multiple acquisition and attrition sequences (e.g., relearning, restoration). A number of these aspects of development will be addressed in section 5.3 (see Hypothesis 12 in section 2.5).

A fourth implication for this study is that other details of the developmental course require investigation: which specific language items and which learners will benefit from which types of instruction. In other words, specific language items may have different kinds of learning and forgetting curves. For this study, I have selected a language element and a set of learners that may be particularly revealing because of the acknowledged difficulty of English article use for learners whose first language does not have articles (Celce-Murcia & Larsen-Freeman, 1983; Master, 1987). In sum, this consideration of the temporal nature of the developmental course of instructed SLA indicates that instructional studies need to be carefully designed in order to find lasting effects, and that permanent effects may not be as pedagogically important as has previously been believed.

2.4 The L2 Development of Form-function Mappings: Interacting Cognitive and Linguistic Principles

This section integrates the issues of form-function mappings, variability and temporal development, attempting to develop an interacting set of cognitive and linguistic principles that account for certain aspects of the nature of instructed SLA. As an empirical study, the data collected for this investigation will lead to the *description* of the conditions under which learners supply the article *the* in a target-like fashion. Thus, on a spoken task *the* may be supplied more often in non-subject syntactic positions than in subject positions, and more often when the noun it precedes (and specifies, cf. Abney, 1987) was not mentioned in the previous clause. In addition, the study seeks to describe the changes in this distribution of *the* that occur after learners receive instruction, assessing whether instruction alters interlanguage knowledge and leads to more target-like use. However, this study seeks to go beyond description and therefore also attempts to *explain* why articles are supplied in these patterns and why instruction has effects on these patterns. By considering data in relation to previously motivated theoretical explanations, the descriptive results of case studies may be generalized beyond the individuals studied to a theory of instructed SLA (cf. Yin, 1989). Although this study will be explained within a specific theoretical framework, the empirical results may be subject to alternative explanations.

The theoretical framework that is adopted in this study assumes that there are interacting cognitive and linguistic (functional and discursal) principles that result in specific patterns of interlanguage use and development (cf. Ellis, 1994). The study is concerned with how L2 learners establish and use form-function mappings, that is, with the factors that result in learners using certain linguistic structures to use in order to express meaning. In languages and in interlanguages, the mappings between form and function may not be one-to-one: one form may serve more than one function, or one function may be expressed by more than one form or structure. Language learners will need to determine the appropriate form-function mappings of the target language and will need to resolve non-target-like mappings that develop during the learning process. The approach in this study is consistent with the set of operating principles developed in Andersen (e.g., 1990a). Andersen (1990a) provides a number of principles, derived primarily from functionalist and cognitive approaches to language learning (e.g., Slobin, 1985), which are

hypothesized to guide second language learning and use. In addition, the approach in this study is consistent with the cognitive processes developed in McLaughlin (e.g., 1987, 1990). McLaughlin suggests that automatization and restructuring are two particularly important processes that occur in all types of learning, including second language learning. This study combines these two approaches, considering how appropriate form-function mappings are established or achieved, potentially through processes such as automatization and restructuring. As discussed in detail by McLaughlin (1990), cognitive processes such as automatization and restructuring may be initiated or affected by instruction and practice.

In order to understand the development of the article *the* and the effects of instruction on that development, the approach in this study considers two different types of form-function mappings: (a) redundancy mapping, the interaction of multiple structures that express the same function, and (b) polyfunctional mapping, the interaction of multiple functions expressed by the same form. These two types of form-function mappings are illustrated in Table 2.3, using the linguistic information that was presented in sections 2.1.3.1 and 2.1.3.2.

Table 2.3
Possible Interlanguage Form-function Mappings of Articles

Tracking Function	Structure					Mapping Relation
	a	the	Ø	Sub	NSub	
A. New NP (unpredictable)	√ 1				√ 2	Redundancy
B. Assumed known NP		√ 1		√ 2		Redundancy
1. old referent (unpredictable)		√				Poly- func- tional
2. recent referent (unpredictable)		√ 2				
3. current referent (predictable)			√			
Mapping Relation =						

Note. Sub = Subject position; NSub = Non-subject positions. The numerals indicate the first and second structures (numbered horizontally) or functions (numbered vertically) in the redundancy and polyfunctional mappings.

As Table 2.3 indicates, two structures, *the* and subject position, may be used to indicate that the referent of an NP is assumed known to the listener/reader, constituting a redundancy mapping. Similarly, two other structures, *a* and non-subject positions, may be used to indicate that the referent of an NP is new to the listener/reader, also constituting a redundancy mapping. In addition, Table 2.3 also indicates that *the* may serve two functions, indicating that the referent of an NP is assumed known to the listener/reader and also indicating that a known referent is relatively unpredictable (recent or old), constituting a polyfunctional mapping

Although a redundancy mapping may have many possible variations, the type of interaction that will be examined in this study is defined in 12) below.

12) redundancy mapping: the interaction in interlanguage of multiple forms or devices that express the same function and that may appear simultaneously in an utterance, sentence or word.

In this type of form-function mapping, the forms are redundant: they are not mutually exclusive of each other and are not competing in a learner's interlanguage system to express the same function (cf. Ellis, 1994, p. 366). In other words, the two forms in this type of mapping do not occupy the same position in a sentence or word and therefore are neither in free variation nor in complementary distribution (to borrow terminology from phonology). Thus, this type of mapping is not concerned with the presence in interlanguage of two forms that express the same function, such as *no* and *don't* (Ellis, 1985), or *the* and *this*, or *a* and *one* (e.g., Mellow & Fuller, 1995). The redundant structures may be developmental (i.e., non-target-like) or may be target-like, because languages are very often redundant, providing more than one means of expressing certain types of meaning.

Although the study reported here is, in a certain general sense, concerned with two non-redundant forms, *the* and \emptyset , which are competing to indicate that the referent of a noun is already known to the listener,¹¹ one of the two major linguistic analyses in this study is concerned with a redundancy mapping: the use of both *the* and subject position to indicate that the referent of a noun

¹¹ For example, \emptyset could stand in opposition to *a* in an interlanguage and therefore effectively mark whether the referent is known to the listener.

phrase is assumed known to the hearer (cf. section 2.1.3.1, Table 2.3, and Principle 1 below). In this study, the analysis of redundancy mappings is informed primarily by the theoretical principles of cognition and language developed in Mellow and Cumming (1994), discussed as Principle 3 below, as well as by the cognitive process of automatization (e.g., McLaughlin, 1990), discussed as Principle 4 below.

The second major type of form-function mapping, polyfunctional mappings, is the interaction of multiple functions expressed by the same form. Although this mapping may have many possible variations, the type of interaction that will be examined in this study is defined in 13) below.

13) polyfunctional mapping: the interaction in interlanguage of multiple functions that are expressed by the same form and that may simultaneously affect the use of that form.

This mapping relation can be considered to be the intersection of two or more functional systems, with a subset of the structures used to express one functional system overlapping with a subset of the structures of another functional system. Studies of determiner systems in interlanguage have often considered and analyzed complex sets of this type of mapping. For example, Huebner (1979, 1983a, 1983b) examined the interlanguage development of English article use, considering how articles were used to express two semantic functions: (a) whether the information expressed by a noun phrase is assumed to be known to the hearer, and (b) whether the noun phrase has specific or generic reference (cf. section 2.1.2). Similarly, Pfaff (1987) considered the manner in which German interlanguage determiners were used to express a variety of functions.

The second of the two major linguistic analyses in this study is concerned with a polyfunctional mapping: the use of *the* to indicate referential distance (that the referent was not mentioned recently in the discourse) and to indicate that the referent is assumed known to the hearer (cf. section 2.1.3.2 and Table 2.3). In this study, the analysis of the polyfunctional mapping is informed by the theoretical principles of cognition and language that will be discussed as Principle 2 below.

As can be seen from the preceding discussion, the use of any linguistic element will be affected by a large number of factors: One of the most inescapable facts of interlanguage development is that it is particularly complex and variable (cf. Gregg, 1984; Larsen-Freeman & Long, 1991, p. 14). In this study, the suppliance of *the* in interlanguage is being considered only in reference to two functions that it serves and to its redundant status for expressing one of those functions. Clearly, there will be many other factors involved in the acquisition and use of *the*, including, for example, the ways in which the overall article system is acquired, with *the* used in various ways with generic reference nouns and with proper nouns and in opposition to *a* and \emptyset (cf. section 2.1.2). The L2 use of *the* can also be considered in terms of the larger system of prenominal elements (e.g., demonstrative pronouns, possessive pronouns, *one* + noun) and in terms of the larger system of the tracking of NP reference (e.g., the use of pronouns and null forms/ellipsis) (cf. Chaudron & Parker, 1990; Mellow & Fuller, 1995). In addition, a multitude of other factors will affect L2 development and use, with the strength of these factors being influenced by L1 transfer and the attentional requirements of specific tasks, among other considerations (cf. Hatch, 1983; VanPatten, 1984, 1994). Thus, like Mellow and Cumming (1994) and Mellow (in press), this study adopts an interacting processes model, maintaining the view that no single factor nor even a small number of factors will account for the developmental progression of interlanguage elements (cf. Long & Sato, 1984; Young, 1989; inter alia). The multiplicity of factors involved in language acquisition may result in relatively few linear developmental sequences and is one of the reasons why the effects of instruction are difficult to determine. The empirical implication of this complexity is that interlanguage data will exhibit considerable complexity and variability, and that it may be difficult to isolate patterns which, when quantified, achieve levels of significance using inferential statistical procedures. In spite of this assumption of complexity, this study nonetheless makes the assumption that certain strong patterns will emerge and that these patterns can be verified and replicated (cf. the first assumption in section 1.3).

As noted above, the approach in this study is consistent with the set of functional and cognitive principles developed in Andersen (1990a). In particular, one of Andersen's principles is particularly relevant to the approach adopted in this study. This principle, the Multifunctionality Principle, is stated as follows (Andersen, 1990a, p. 53):

(a) Where there is clear evidence in the input that more than one form marks the meaning conveyed by only one form in the interlanguage, try to discover the distribution and additional meaning (if any) of the new form. (b) Where there is evidence in the input that an interlanguage form conveys only one of the meanings that the same form has in the input, try to discover the additional meanings of the form in the input.

The two sub-parts of Andersen's principle refer to the two types of form-function mappings defined above. Andersen's sub-principle (a) includes redundancy mappings. In addition, sub-principle (a) could also refer to elements that do *not* appear simultaneously in an utterance, sentence or word, that is, forms that appear in free variation or complementary distribution. Andersen's sub-principle (b) refers to polyfunctional mappings. Importantly, Andersen's principles are concerned with the discovery or acquisition of these mappings: the principles which lead learners to *restructure* their linguistic knowledge in order to acquire these mappings. In contrast, as will become clear in the following discussion of six linguistic and cognitive principles, this study is primarily concerned with the *use* of the form-function mappings after they have been added to the learner's knowledge and with the *automatization* of these mappings.

In order to interpret the data from this study (i.e., patterns in the distribution of *the* suppliance), it will be assumed that redundancy and polyfunctional mappings interact in a specific manner with the following six hypothesized linguistic and cognitive principles, and that this interaction can account for certain aspects of the nature of interlanguage use and development.

These six hypothesized principles (Ps) are:

P1) An interlanguage system will include a topic-comment system, especially if a learner's first language has a topic-comment system, resulting in a redundancy mapping for *the* in subject (or first NP) position.

P2) In those contexts where an element is polyfunctional, expressing more than one function, that element will be supplied more consistently than when it only expresses one function.

P3) Redundant grammatical elements will be efficiently omitted in tasks which make high demands on attentional resources (including certain types of spoken tasks), but will be supplied more consistently, due to priming, in tasks which do not make high demands on attentional resources (such as writing) (Mellow and Cumming, 1994).

P4) As a learner's knowledge of a redundant form-function mapping becomes increasingly automatized, the efficient elimination of the redundant grammatical form will no longer occur in tasks that have high attentional demands.

P5) Interlanguage behavior on spoken tasks (or for speech styles in which relatively little attention is given to the monitoring of speech and specific language elements) will more clearly exhibit systematic interlanguage patterns and form-function mappings than will written tasks (or for styles in which greater attention is given to the monitoring of speech) (Labov, 1972, p. 208; Tarone, 1988).

P6) Systematic interlanguage patterns that appear when learners have low or moderate proficiency with certain form-function mappings may be obscured by very high competence with those elements, that is, by a ceiling effect.

The first hypothesized principle proposes that learners, especially the Japanese learners in this study, will have a topic-comment interlanguage system. This hypothesis was discussed in detail in section 2.1.3.1. If the learners in this study do have a topic-comment system, then it is expected that this will lead to a redundancy mapping for *the* in subject (or first NP) position, because both *the* and the first NP position would indicate that the referent is assumed known to the hearer (cf. Table 2.3). In addition, although not investigated in this study, the presence of a topic-comment system would imply that new referents do not normally appear in subject position, but appear more often (and therefore more redundantly) in non-subject positions (cf. Table 2.3).

The second hypothesized principle proposes that an element will be supplied relatively more consistently when it is polyfunctional than when it only expresses one function. This assumption would appear to be consistent with models of cognitive processing, such as connectionist models, which suggest that when a form serves two functions the inputs will conspire to strengthen patterns of neurological connectivity for that form (cf. Hinton, McClelland, & Rumelhart, 1986, pp. 81-82). Assuming that Givón's Quantity Universal is correct (section 2.1.3.2), this principle predicts that *the* will be supplied more consistently when it indicates that a

referent is both assumed known to the hearer and has greater referential distance (i.e., the referent was not mentioned in the previous clause).

The third hypothesized principle proposes that if the model of language processing outlined in Mellow and Cumming (1994) is correct, then redundant grammatical elements (two forms that express the same function) will be efficiently eliminated in tasks which make high demands on attentional resources (especially aural comprehension, but also in certain types of spoken tasks), but will be supplied more consistently, due to priming, in tasks which do not make high demands on attentional resources (such as composition) (cf. VanPatten, 1994). Drawing upon theoretical models of language processing cognition, Mellow and Cumming (1994, pp. 446-447), proposed the Priming Hypothesis:

For language tasks that involve relatively minor restrictions on processing resources, concorded grammatical morphemes are processed more easily and used with greater accuracy than non-concorded grammatical morphemes because the preceding concurring elements prime and facilitate the processing of the grammatical morphemes.

Mellow and Cumming found evidence in SLA data which suggested that priming will be a significant factor only in those situations, primarily language production or output tasks, which permit an L2 speaker to control and slow the pace of language processing and hence to increase the availability of attentional resources. In addition, Mellow and Cumming found evidence which suggested that during many input tasks or rapid conversation, which require that limited attentional resources be directed to the comprehension and expression of meaning, the priming of concorded affixes was not a significant factor. The data therefore also supported an efficiency hypothesis: In tasks where efficiency is most required and attentional resources are very limited, the efficient omission of concorded affixes is expected to be a significant factor.

The priming and the efficiency hypotheses were proposed and examined in reference to concorded grammatical morphemes: instances in which the grammatical information encoded in a grammatical morpheme was also provided by a preceding lexical item, for example, for number: *the TWO book+S*. The current study extends those proposals and considers the more general phenomenon of redundancy, hypothesizing that priming and efficiency occur in reference to all

redundant marking of grammatical information, and are not restricted to concorded grammatical affixes (cf. Mellow & Cumming, 1994, p. 466). The facilitation provided by redundancy, the multiple activation of the cognitive representation of one meaning by redundant forms, is therefore similar to the facilitation provided by polyfunctionality, the multiple activation of the cognitive representation of one form for different functions. The important difference between these hypothesized cognitive processes is that redundant mappings will result in facilitation only under certain task conditions, those with abundant attentional resources, and that the efficient omission of redundant elements will be a stronger factor when tasks do not permit the L2 speaker to control and slow the pace of language processing. In contrast, the facilitation provided by polyfunctionality will be expected to be manifested under all task conditions.

Two empirical predictions arise from the third hypothesized principle, in conjunction with the linguistic assumptions in Principle 1. First, in a written task and due to priming, *the* will be supplied more consistently when it is in subject (first NP) position than when it is in non-subject (postverbal NP) positions. Second, in a spoken task in which the speaker must focus carefully on the content being conveyed and due to efficiency, *the* will be supplied less consistently when it is in subject (first NP) position than when it is in non-subject (postverbal NP) positions. These predictions indicate the importance of task variation, discussed in section 2.2, and necessitate a careful selection of tasks, discussed in section 4.1.3.

The predictions of this processing account, with respect to the redundancy of articles, has been confirmed in studies of SLA (cf. Duff, 1985, pp. 3, 54, 70; Huebner, Perdue & Carroll, 1988, p. 135). The efficient omission of definite articles in subject position in spoken data was found for the Hmong speaking English learner reported in Huebner (1979, p. 26) and the Punjabi speaking English learner reported in Huebner, Perdue and Carroll (1988, pp. 53, 58). In addition, the more frequent suppliance of indefinite articles with new referents that are introduced in the subject position (first NP) rather than non-subject positions has also been reported by SLA researchers (e.g., Huebner, Perdue, & Carroll, 1988, pp. 47, 135).

The fourth hypothesized principle proposes that as a learner's knowledge of a redundant form-function mapping becomes increasingly automatized, the efficient elimination of redundant grammatical forms will no longer occur in tasks that have high attentional demands, such as a spoken task in which the speaker must focus carefully on the content being conveyed. By considering the effects of automatization, one of the two cognitive learning principles discussed by McLaughlin (1990), this principle is able to make predictions about the nature of L2 development. The central aspect of this principle is that the use of automatized knowledge requires fewer attentional resources than does non-automatized or less automatized knowledge. As a result, automatized form-function mappings will not be competing for a speaker's limited attentional resources, and therefore efficiency will not be a strong factor in the use of that language element. Automatization of specific redundant form-function mappings is predicted to occur as a learner develops due to naturalistic (uninstructed) language acquisition, but may be particularly facilitated by instruction and practice (e.g., McLaughlin, 1990). The specific empirical prediction of this principle is that the non-subject > subject asymmetry in the suppliance of *the*, due to the efficient omission of *the* in subject position in an attention-demanding spoken task, will be reduced or eliminated as the redundancy mapping becomes increasingly automatized. An additional possibility is that, at very high levels of automatization, priming will occur in an attention-demanding spoken task.

Automatization is also expected to be revealed in the use of other form-function mappings and on other tasks. However, this manifestation would be expected in the form of a general increase in suppliance of *the*, rather than in terms of changes in interlanguage patterns.

Specifically, the priming of redundant elements, revealed as a subject > non-subject asymmetry in the suppliance of *the* in a written task, would not be expected to be affected by automatization because it is predicted to occur in a task situation where there is relatively little competition for

attentional resources. In addition, the facilitation resulting from polyfunctionality would not be predicted to be affected because it is not dependent on the availability of attentional resources.¹²

The nature of automatization may also vary in certain ways that are open to empirical verification. For example, automatization could be gradual or immediate, an issue discussed in section 2.3 and explored in section 5.3. In addition, automatization is not the only cognitive process that may result in changes to L2 knowledge. Other potential processes include: 1) the addition of new knowledge, potentially resulting in restructuring; 2) the proceduralization of declarative knowledge; and 3) the loss of knowledge, especially if that knowledge had not been automatized or proceduralized. Some of these possibilities will be investigated in sections 5.1 and 5.3. As discussed in Chapter 1, findings which indicate whether instruction can result in the automatization of L2 knowledge are important both for a theory of SLA and for the pedagogical assessment of form-focused instruction.

The fifth hypothesized principle proposes that interlanguage behavior on a spoken task (or in the vernacular style) will more clearly exhibit systematic interlanguage patterns than will writing and more monitored tasks (or in the careful styles) (cf. Tarone, 1988, p. 40). This may be the case because careful or monitored tasks involve a greater use of non-automatized, conscious, and potentially metalinguistic grammatical knowledge. This knowledge may either be inaccurate or inconsistently applied (cf. Krashen, 1982), or may lead to a more target-like use that obscures, overlays, or drowns out any systematic interlanguage patterns that are part of non-conscious competence.

Similar to the fifth principle, the sixth and final hypothesized principle proposes that the systematic interlanguage patterns that may appear when learners have low or moderate proficiency with certain form-function mappings may be obscured by very high competence with those mappings, that is, by a ceiling effect. In other words, when learners reach their maximum ability

¹² In addition, improvements in language use on spoken tasks with limited availability of attentional resources will be especially suggestive that automatization has occurred, since improvements on the written task may also be explained by changes in non-automatic knowledge (cf. Krashen, 1993; and sections 1.3 and 2.3).

with a form-function mapping, any sub-divisions or patterns within the use of that mapping may no longer be manifested (cf. the results on the cloze task discussed in section 5.1). For example, when learners supply a high percentage of *the* in both subject and non-subject positions, any priming or efficiency effects may be obscured. Similarly, when learners supply a high percentage of *the*, the facilitating effects of polyfunctionality may be obscured, a result revealed in the data reported by Chaudron and Parker (1990, p. 57), with the higher provision of bare nouns for current topics not appearing in the high proficiency group.

It is difficult to determine *a priori* the percentage level of suppliance or target-like use that might be a ceiling for any individual learner because the use of that mapping may be affected by many other factors, such as a difficulty in producing certain consonant clusters or the presence of a non-target-like lexical representations of certain word (e.g., whether nouns are mass or count) (cf. section 2.3.1, footnote 10), and because the use of interval scales and percentages may not accurately reflect actual interlanguage development (cf. Lantolf & Frawley, 1988, discussed in section 3.3). This sixth principle is related to the previous principle in that a systematic interlanguage pattern may be obscured in a written task either by the use of conscious knowledge or by the achievement of a ceiling or both, and it may be difficult to distinguish between the two effects. However, the drowning out effect of conscious knowledge is hypothesized to occur only on certain tasks, such as in writing, and the drowning out effect of reaching a ceiling effect is hypothesized to occur only at a relatively high level of competence.

In sum, this section has discussed six linguistic and cognitive principles which inform this study of instructed second language acquisition. Although the data itself will always be subject to differing interpretations, results which are consistent with the patterns predicted by these hypotheses will have external validity, generalizing not to a larger population but to a developing and increasingly coherent theory of instructed second language acquisition (cf. section 3.3.2). The principles that have been proposed have drawn upon a number of fundamental, and relatively uncontroversial linguistic and cognitive processes (cf. Mellow, in press), including form-function

mappings, automatization, restructuring, transfer, attention, efficiency (omission or simplification), priming, as well as conscious and unconscious knowledge of language use.

2.5 *Research Hypotheses*

The six cognitive and linguistic principles presented in the previous section motivate a series of twelve empirical hypotheses or predictions regarding the effects of instruction on the use of form-function mappings in interlanguage. These predictions are formulated in reference to the method and analyses justified and described in Chapters 3 and 4. In the study, two groups of four learners, one instructed and one uninstructed, provided language samples on three tasks, spoken, written and cloze retellings of a fable. The samples were provided during a six month period that was divided into two phases: a preintervention (PreI) phase and a postintervention (PostI) phase. The data from those samples was considered in terms of the following five dependent variables: (1) *the* suppliance for referential definites (+SR, +HK; abbreviated Def), (2) *the* suppliance in subject position (abbreviated Sub), (3) *the* suppliance in non-subject position (abbreviated NSub), (4) *the* suppliance for recent referents (abbreviated Rec), and (5) *the* suppliance for current referents (abbreviated Cur). The first dependent variable is assessed on all three tasks and the other four are assessed on only the spoken and written fable retellings. For each of these dependent variables, the scores from the data collection sessions are summarized in two ways, (1) in terms of central tendency, using group mean phase scores and inferential statistical tests; and (2) in terms of trend, using group phase regression lines and line graphs of individual development.

The twelve empirical hypotheses are sub-divided into four categories. The first category of hypotheses is concerned with the nature of interlanguage use during the preintervention (PreI) period, examining whether interlanguage use is characterized by variation according to task and linguistic context. Determining the interlanguage behavior of all of the learners during the baseline period is particularly important because it is only with a clear understanding of the baseline behavior of both groups that one can precisely determine the nature of any possible changes after

intervention, thereby noting group changes and differences. Because the effects of instruction may be complex, they can only be assessed in light of a clear understanding of previous behavior.

The first hypothesis, regarding task variation during the PreI phase, states that *the* suppliance for referential definites on the cloze task will be more target-like than *the* suppliance on the written task, and *the* suppliance for referential definites on the written task will be more target-like than *the* suppliance on the spoken task, leading to the following hierarchy of scores: cloze task > written task > spoken task. The prediction for these data is based on the hypothesis that greater attention to form leads to more target-like use in interlanguage (cf. the discussion of variability in section 2.2)

The second hypothesis, regarding the effects of syntactic position on the spoken task, states that there will be a NSub > Sub asymmetry in the spoken data (due to efficient omission of *the* in the subject position): *the* will be supplied more often in non-subject positions than in subject position. The prediction is based upon previous empirical findings and upon the principles of redundancy outlined in Principle 3.

The third hypothesis, regarding the effects of syntactic position on the written task, states that there will be a Sub > NSub asymmetry in the written data (due to priming of *the* in subject position): *the* will be supplied more often in subject position than in non-subject positions. The prediction is based upon previous empirical findings and upon the principles of redundancy outlined in Principle 3.

The fourth hypothesis, regarding the effects of referential distance on the spoken task, states that there will be a Rec > Cur asymmetry in the spoken data: *the* will be supplied more often for recent referents than for current referents. The prediction is based upon previous empirical findings, upon Givon's (1983) Quantity Universal, and upon the principles of polyfunctionality outlined in Principle 2.

The fifth hypothesis, regarding the effects of referential distance on the written task, states that there will be a Rec > Cur asymmetry in the written data: *the* will be supplied more often for recent NPs than for current NPs. However, these effects may be moderated or obscured by an

additional factor: The application of metalinguistic knowledge may affect *the* suppliance on the written task. These predictions are based upon Givon's (1983) Quantity Universal, upon the principles of polyfunctionality outlined in Principle 2, and upon the principles governing the application of metalinguistic knowledge outlined in Principle 5.

The second category of hypotheses is concerned with the effects of the instruction on the five dependent variables. The sixth hypothesis, regarding group differences in gains in mean phase scores, states that with respect to the five dependent variables on the written, spoken, and cloze tasks, the instructed group will exhibit greater gains in mean phase scores from the PreI phase to the PostI phase than will the uninstructed group. The group differences are expected as a result of effective instructional practice activities that lead to increased automatization of these form-function mappings, although these effects may be moderated by the fact that the two groups are not equivalent (cf. section 5.1.3). These predictions are based upon the principles of practice and automatization outlined by McLaughlin (1990) (cf. section 1.3).

The seventh hypothesis can be divided into two parts. Hypothesis 7a, regarding changes in developmental trend in the instructed group, states that with respect to the five dependent variables on the written, spoken, and cloze tasks, the instructed group will exhibit continued or improved developmental trends from the PreI phase to the PostI phase. By using the baseline trend from the PreI phase, the treatment group acts as their own controls. The continued or improved development is expected as a result of effective instructional practice activities that lead to increased automatization of these form-function mappings. These predictions are based upon the principles of practice and automatization outlined by McLaughlin (1990) (cf. section 1.3).

Hypothesis 7b, regarding group differences in changes in developmental trend, states that with respect to the five dependent variables on the written, spoken, and cloze tasks, the instructed group will exhibit a greater number of positive changes in developmental trend from the PreI phase to the PostI phase than will the uninstructed group. The group differences are expected as a result of effective instructional practice activities that lead to increased automatization of these form-function mappings, although these effects may be moderated by the fact that the two groups are not

equivalent (cf. section 5.1.3). These predictions are based upon the principles of practice and automatization outlined by McLaughlin (1990) (cf. section 1.3).

The third category of hypotheses is concerned with the effects of the instruction, in terms of the interaction between task, linguistic context and instruction. The eighth prediction can be divided into two parts. Hypothesis 8a, regarding changes in the efficient deletion of redundant elements on the spoken task for the instructed group, states that with respect to *the* suppliance in subject and non-subject positions on the spoken task, the instructed group will exhibit a reduction in the NSub > Sub asymmetry in the Post I phase. The changes will be assessed using group mean phase scores as well as the levels and slopes of the group phase regression lines. By using the baseline trend from the PreI phase, the treatment group acts as their own controls. The loss of the NSub > Sub asymmetry is predicted because instruction is expected to result in the automatization of this procedural interlanguage knowledge, thereby reducing the subtractive effects of the efficient deletion of the redundant use of *the* in subject position. These predictions are based upon the principles of redundancy outlined in Principle 3 and upon the principle of automatization outlined in Principle 4.

Hypothesis 8b, regarding group differences in changes in the efficient deletion of redundant elements on the spoken task, states that with respect to *the* suppliance in subject and non-subject positions on the spoken task, the instructed group will exhibit a greater reduction in the NSub > Sub asymmetry in the PostI phase than will the uninstructed group. The changes will be assessed using group mean phase scores as well as the levels and slopes of the group phase regression lines. The group differences are expected as a result of effective instructional practice activities that lead to increased automatization of this form-function mapping, with this automatization greater than that resulting only from exposure to English. These group effects may be moderated by the fact that the two groups are not equivalent (cf. section 5.1.3). These predictions are based upon the principles of redundancy outlined in Principle 3 and upon the principle of automatization outlined in Principle 4.

The ninth hypothesis, regarding changes in the priming effect of redundant elements on the written task for both groups, states that with respect to *the* suppliance in subject and non-subject positions on the written task, both groups will exhibit a continued Sub > NSub asymmetry in the PostI phase. However, if one or both groups achieve a ceiling (near 100% suppliance), then the facilitating effects of priming may be obscured, with suppliance being equal in the two positions. The changes will be assessed using group mean phase scores as well as the levels and slopes of the group phase regression lines. By using the baseline trends from the PreI phase, the groups act as their own controls and are also compared to each other. The Sub > NSub asymmetry is not predicted to disappear after instruction because increased automatization will not affect the facilitating effect of priming on tasks where relatively abundant attentional resources are available, as in the written task. Because priming is a cognitive process that occurs whenever redundancy and sufficient attentional resources are present, priming effects are predicted to disappear only if the redundancy was eliminated, that is, if the interlanguage was restructured, removing the topic-comment system. These predictions are based upon the principles of redundancy outlined in Principle 3 and upon the ceiling effect outlined in Principle 6.

The tenth hypothesis, regarding changes in the facilitating effect of polyfunctional elements on the spoken task for both groups, states that with respect to *the* suppliance for current and recent referents on the spoken task, both groups will exhibit a continued Rec > Cur asymmetry in the PostI phase. However, if one or both groups achieve a ceiling (near 100% suppliance), then the facilitating effects of polyfunctionality may be obscured, with suppliance being equal for both types of referents. The changes will be assessed using group mean phase scores as well as the levels and slopes of the group phase regression lines (see section 4.2.2). By using the baseline trends from the PreI phase, the groups act as their own controls and are also compared to each other. The Rec > Cur asymmetry is not predicted to disappear after instruction because increased automatization will not affect the facilitating effect of polyfunctionality. The facilitation provided by polyfunctionality is a cognitive process that occurs whenever polyfunctionality is present, independent of the attentional demands of tasks and the degree of automatization of the

interlanguage knowledge. These predictions are based upon the principle of polyfunctionality outlined in Principle 2 and upon the ceiling effect outlined in Principle 6.

The eleventh hypothesis, regarding changes in the facilitating effect of polyfunctional elements on the written task for both groups, states that with respect to *the* suppliance for current and recent referents on the written task, both groups will exhibit a continued Rec > Cur asymmetry in the PostI phase. However, if one or both groups achieve a ceiling (near 100% suppliance), then the facilitating effects of polyfunctionality may be obscured, with suppliance being equal for both types of referents. In addition, these polyfunctionality effects may be obscured on the written task by the use of metalinguistic knowledge or other conscious strategies. The changes will be assessed using group mean phase scores as well as the levels and slopes of the group phase regression lines (see section 4.2.2). By using the baseline trends from the PreI phase, the groups act as their own controls and are also compared to each other. As noted in Prediction 10, the Rec > Cur asymmetry is not predicted to disappear after instruction because increased automatization will not affect the facilitating effect of polyfunctionality. These predictions are based upon the principle of polyfunctionality outlined in Principle 2, upon the principle governing the application of metalinguistic knowledge outlined in Principle 5, and upon the ceiling effect outlined in Principle 6.

The fourth category of hypotheses is concerned with the effect of instruction, for specific individuals, on the developmental course of SLA, including variations such as immediate or gradual improvement and U-shaped development. The twelfth hypothesis, regarding the patterns of individual learners with respect to the developmental course of instructed SLA, states that with respect to the five dependent variables on the written, spoken, and cloze tasks, each instructed and uninstructed learner will exhibit unique patterns of development, possibly including gradual improvement, immediate improvement, no improvement, and delayed attrition. These predictions are based upon the nature of developmental course discussed in section 2.3 and the concerns for individual variation discussed in section 3.3.2.

Chapter 3

Issues of Research Design and Measurement in Instructional Studies

The first two chapters have indicated the scope of the problem that is investigated here, in terms of pedagogical issues (Chapter 1) and in terms of the nature of interlanguage development (Chapter 2). However, establishing a causal relationship between sessions of instruction and changes in the interlanguage system presents a very difficult challenge. This chapter begins with a review of previous instructional studies, focusing on methodological issues, and then addresses the solutions adopted in this study.

3.1 A Methodological Review of Instructional Studies

The effects of instruction on L2 development have been extensively studied during the last several decades. A large number of SLA researchers have provided reviews of instructional studies, including Cadierno (1992), Ellis (1990a; 1994, pp. 611- 663), Harley (1989b), Larsen-Freeman and Long (1991, pp. 299-331), Long (1983, 1988), Mellow (1992), Pienemann (1985), Spolsky (1989, pp. 187-201), and VanPatten (1988). Because of the complexity of instructional effects, introduced in Chapter 1, these many studies and these many reviews have not yet found conclusive evidence that form-focused instruction helps L2 learners in important or meaningful ways (cf. Ellis, 1988, p. 26; 1994; Harley, 1989b, pp. 170, 172; Krashen, 1992, 1993; Long, 1983, p. 380; Pienemann, 1985, pp. 30, 33; VanPatten, 1988, pp. 249-253). Although a review of instructional studies is provided here, its purpose is not to differentiate between the effects of varying types or amounts of instruction, or to consider the theoretical implications of the results (e.g., with respect to parameter setting, implicational hierarchies, developmental sequences, or learner orientation and social distance). Instead, the purpose of this review is to consider the methodological features of those studies, in terms of design (the delay until post-tests) and

measurement instruments (the use of tasks that encourage varying amounts of attention to the element being investigated).

The review is representative rather than exhaustive or comprehensive (see Ellis, 1994 for a more comprehensive review). In addition, the review includes only a subset of L2 training studies. Only those studies that have examined the development of specific grammatical (morpho-syntactic) elements of language use are included. Furthermore, the review does not consider studies of instruction that compared large units of instruction (e.g., an L2 class for an entire term) as measured on relatively general measures of language ability (e.g., a TOEFL score; cf. section 2.1) because these units of analysis are believed to be too general and therefore have led to ambiguous or inconsistent results because distinct aspects of language or instruction, or both, may have overlapped (Chaudron, 1988, pp. 165, 166; Doughty, 1991; Ellis, 1988, pp. 26-27, 1990a, p. 146; Harley, 1989b, p. 170; Larsen-Freeman & Long, 1991, p. 322; Long, 1983, p. 380; Mellow, 1992; Sharwood Smith, 1981, p. 167; VanPatten, 1988, pp. 245, 252, 255; cf. Andersen, 1982; Lambert & Moore, 1986; and Weltens, 1987, p. 31). Instead, the studies reviewed are a subset of those that are precise and constrained in terms of the scope of the unit of instruction and the aspect of language that is investigated (cf. section 2.1). Instructional studies were included if they provided specific form-focused instruction (presentation, practice, or correction; cf. consciousness raising and input enhancement: Sharwood Smith, 1981, 1991, 1993) about specific aspects of language and then measured learning on instruments which were sensitive to use of that specific aspect of language. One must be cautious in interpreting the value of linguistically constrained studies because, as discussed in Chapter 2, greater specificity may clarify or disambiguate aspects of language use, but may also may oversimplify the complex interactive nature of linguistic systems.

The first methodological factor that was considered in this review, delay until post-tests, was divided into the three categories in 14), 15) and 16) below.

- 14) Short-term - less than or equal to 3 days;
- 15) Mid-term - between 6 days and 5 weeks;
- 16) Long-term - more than 9 weeks.

The lack of contiguity between the categories (i.e., the jump from 3 to 6 days, the jump from 5 to 9 weeks) reflects the actual time frame involved in the studies reviewed rather than any pre-determined principled reason. The second methodological factor that was considered, measurement instrument, was divided into the three categories in 17), 18) and 19) below, using inferences based upon the descriptions of the measurement instruments:

17) monitored instruments - learners probably focused their attention on specific language items and probably had time to draw upon their metalinguistic knowledge; example instruments: grammaticality judgements, preference tasks;

18) partially monitored instruments - learners may have focused their attention on specific language items but may not have had time to draw upon their metalinguistic knowledge; example instruments: translation, compositions, specific oral questions; and

19) unmonitored instruments - learners probably did not focus their attention on specific language items and probably did not have time to draw upon their metalinguistic knowledge; example instruments: picture descriptions.

Nineteen instructional studies were reviewed with respect to these categories and the effect of instruction (positive effect or no effect) that was reported by the authors. The studies involved between one and three post-test sessions, used as many as three different instruments, and may have reported results on more than one type of instruction. The results of a total of 60 post-tests are therefore reported in Table 3.1.

As indicated in Table 3.1, 100% of those studies that assessed instructional effects on short-term, monitored post-tests reported positive effects (positive effect: Schumann, 1978; Gass, 1982; Eckman, Bell, & Nelson, 1988; Kadia, 1989; Harley, 1989a; White, 1991; Day & Shapson, 1991; White, Spada, Lightbown, & Ranta, 1991). Sixty percent of those studies that assessed instructional effects on mid-term, monitored post-tests reported positive effects (positive effect: White, 1991; Carroll, Roberge, & Swain, 1992; White, Spada, Lightbown, & Ranta, 1991; no effect: Carroll & Swain, 1993; Carroll, Roberge, & Swain, 1992). Forty-three percent of those studies that assessed instructional effects on long-term, monitored post-tests reported positive effects (positive effect: Lightbown, Spada, & Wallace, 1980; Kadia, 1989; Day & Shapson, 1991; no effect: Lightbown, Spada, & Wallace, 1980; Kadia, 1989; Harley, 1989a; White, 1991).

Table 3.1
Number of Instructional Studies Exhibiting a Positive Effect or No Effect Due to Instruction, According to Delay Until Posttest and Type of Test

<u>Instrument</u>	<u>Delay until post-test</u>		
	<u>Short -term</u>	<u>Mid-term</u>	<u>Long-term</u>
Monitored			
positive effect	8	3	3
no effect	0	2	4
Totals (%)	8/8 (100%)	3/5 (60%)	3/7 (43%)
Partially Monitored			
positive effect	13	8	1
no effect	4	1	1
Totals (%)	13/17 (76%)	8/9 (89%)	1/2 (50%)
Unmonitored			
positive effect	2	2	0
no effect	3	1	4
Totals (%)	2/5 (40%)	2/3 (67%)	0/4 (0%)

Seventy-six percent of those studies that assessed instructional effects on short-term, partially monitored post-tests reported positive effects (positive effect: Lightbown, Spada, & Wallace, 1980; Lalande, 1983; Zobl, 1985; Tomasello & Herron, 1988, 1989 [2 results]; Doughty, 1988, 1991; Carroll & Swain, 1993; Carroll, Roberge, & Swain, 1992; Day & Shapson, 1991; Ross & Berwick, 1991; VanPatten & Cadierno, 1991 [2 results]; White, Spada, Lightbown, & Ranta, 1991; no effect: Harley, 1989a; Ross & Berwick, 1991 [3 results]). Eighty-nine percent of those studies that assessed instructional effects on mid-term, partially monitored post-tests reported positive effects (positive effect: Tomasello & Herron, 1988, 1989 [4 results]; VanPatten & Cadierno, 1991 [3 results]; White, Spada, Lightbown, & Ranta, 1991; no effect: VanPatten & Cadierno, 1991). Fifty percent of those studies that assessed instructional effects on long-term, partially monitored post-tests reported positive effects (positive effect: Day & Shapson, 1991; no effect: Harley, 1989a).

Forty percent of those studies that assessed instructional effects on short-term, unmonitored post-tests reported positive effects (positive effect: Harley, 1989a; White, Spada, Lightbown, & Ranta, 1991; no effect: Schumann, 1978; Ellis, 1984; Day & Shapson, 1991).

Sixty-seven percent of those studies that assessed instructional effects on mid-term, unmonitored post-tests reported positive effects (positive effect: Pienemann, 1984, 1989; White, Spada, Lightbown, & Ranta, 1991; no effect: Pienemann, 1984, 1989). None of those studies that assessed instructional effects on long-term, unmonitored post-tests reported positive effects (no effect: Pienemann, 1984, 1989; Kadia, 1989; Harley, 1989a; Day & Shapson, 1991).

Although Table 3.1 reported the percentage of studies exhibiting positive effects of instruction, these percentages must be interpreted with considerable caution. First, the individual studies may have methodological shortcomings, especially with respect to control of internal validity (cf. Ellis, 1990, pp. 155, 158; Krashen, 1992, 1993; see section 3.2 below). Second, the studies themselves are not strictly comparable, with different designs (e.g., sample size, presence of a control group), different learners (e.g., in terms of age, L1, educational background, proficiency level), different language items, and different kinds of instruction.

Given these limitations, a number of interesting patterns are evident, perhaps suggesting directions for research hypotheses. First, on short term monitored post-tests, *all* forms of instruction were effective. Clearly, instruction is effective. Second, as the test type becomes less monitored, immediate results may become less certain, with 76% of those studies that assessed instructional effects on partially monitored tests reporting positive effects, and only 40% of those studies that assessed instructional effects on unmonitored measures reporting positive effects. Third, long term effects (beyond 8 weeks) were uncommon, and, except for Day and Shapson (1991), were found only on monitored measures. In other words, the developmental course of SLA may be characterized by normal learning and forgetting curves: attrition or loss seems to be inevitable. Perhaps not surprisingly, in a similar review of instructional studies, Mellow (1992) found that studies utilizing more extensive instruction (extending for a period of at least 2 weeks, and totaling at least 6 hours) reported more positive effects, especially on unmonitored measures, and especially mid-term results, than did those studies utilizing less extensive instruction (extending for a period of 3 days or less, and totaling 3 hours or less).

As noted in Chapter 1, the studies indicated that L2 instruction almost always results in significant short term improvement if that improvement is measured on tests in which students have time to draw upon their conscious metalinguistic knowledge. However, this result is not surprising because humans clearly have the ability to learn, consciously perform, and then potentially forget various complex cognitive activities. In contrast, the studies reviewed here provided no evidence for the most educationally desirable results, long term effects on unmonitored language use (cf. Krashen, 1992). However, as demonstrated in the discussion of developmental course in section 2.3, long term effects may be very difficult to obtain, particularly if the control group does not stabilize. Therefore, the studies demonstrating mid-term positive effects of instruction on unmonitored measures (Pienemann, 1984, 1989; White, Spada, Lightbown & Ranta, 1991) should be viewed as very positive results. In addition, the number of studies employing mid-term and long-term unmonitored measures is very small, and therefore, a greater number of longitudinal instructional studies is needed to assess this aspect of language development. Thus, although the results of this review are neither exhaustive nor conclusive, they suggest that instruction has positive effects, but that different research methods are needed to provide pedagogically relevant results (cf. section 1.2).

3.2 Experimental Design Issues

Perhaps the most important implication of the review in the previous section is that pedagogically relevant results are obtained when instructional studies utilize mid-term and long-term unmonitored post-tests. In other words, design issues need to be considered so that longitudinal effects can be examined, and issues regarding measurement instruments need to be considered to assess development on tasks that do not involve attention on the element being investigated.

Because educators and theorists are very interested in permanent or lasting changes in unmonitored or automatic language use, as indicated in sections 2.2 and 2.3, it is important to understand why are there so few instructional studies that actually assess these concerns. This

research gap appears to have resulted from the conflict between two approaches or paradigms within applied linguistic research methodology (cf. Mellow, Reeder & Forster, in press). The two contrasting approaches are: quantitative, psychometric, extensive (henceforth Psychometric/Extensive); and qualitative, linguistic, intensive (henceforth Linguistic/Intensive) (cf. Bachman, 1990, pp. 296-334; Cook & Campbell, 1979, 91-94; Johnson, 1992, pp. 31-33; Lantolf & Frawley, 1988; Long & Sato, 1984; Seliger & Shohamy, 1989, pp. 32-40; Yin, 1989, pp. 21, 38-40).

The first approach, Psychometric/Extensive, emphasizes the use of reliable measurement instruments and quasi-experimental control group designs in a search for objective, verifiable knowledge, and often uses inferential statistical tests to assess the behaviors of large, randomly selected samples of learners, divided into treatment and nonequivalent control groups, and to draw inferences about causality that are valid for an entire population. The large sample size, in particular, results in considerable time required to transcribe and analyze the spoken data needed to assess unmonitored language use. As a result, written tests of monitored ability, which may be scored quickly, are used most often in these studies. In addition, long-term study of such a large number of learners is usually very difficult, often due to the temporary nature of the intact classes that are studied, precluding longitudinal studies (cf. Huebner, 1983b, p. 49). As a result, one or a small number of post-tests are usually included. Many of the instructional studies reviewed above in section 3.1 adopted this Psychometric/Extensive approach, often using structural analyses language (e.g., Tomasello & Herron, 1988, 1989; White, 1991).

Although the Psychometric/Extensive type of study emphasizes the determination of causal relationships (internal validity) and the generalizability of results (external validity), these L2 instructional studies often exhibit methodological shortcomings which threaten these types of validity (cf. Beck & Eubank, 1991; Harley, 1989a, p. 335; Nunan, 1991; Tomasello & Herron, 1991). Achieving controlled experimental conditions is very difficult in an applied discipline. As a result, most Extensive instructional studies (i.e., control group designs) are *quasi-experiments* rather than true experiments (e.g., Cook & Campbell, 1979) and hence they may not be able to

clearly establish causality nor support generalizations (for similar discussions, see Parsonson & Baer, 1978, pp. 101-08; Tawney & Gast, 1984, pp. 52-3).

An example of an important limitation of most educational studies is the non-equivalence of groups. This often results because groups are not randomly selected from a defined population. Typically, only volunteers or available groups are used. One serious implication of non-random groups is that causality cannot be clearly established because a group difference other than the provision of treatment may have resulted in the apparent differences. An example of the problem of non-equivalent groups in SLA research can be seen in Doughty (1991). In spite of an attempt to select only learners of a similar ability level (Doughty, 1991, p. 441) and in spite of the random assignment of the 20 learners to three groups (Doughty, 1991, p. 442), the pre-test means of Doughty's three groups were very different. The group pre-test means were 45.8 (treatment group 1), 53.7 (treatment group 2), and 60.4 (control group), compared to post-test scores of 71.1, 80.2, 72.4, respectively (Doughty, 1991, p. 453). Because Doughty's groups were arguably not comparable, claims for causality should be very tenuous. Rather than comparing post-test scores, which showed no between group differences, Doughty compared gain scores (from pre-test to post-test), and did find group differences. These differences in gain scores may be due to the non-equivalence of the groups rather than due to differential effects of treatments. The issue of non-equivalence of groups is discussed further in Chapters 4 and 5.

The second approach to research, Linguistic/Intensive, emphasizes an analysis of the many details of the context in which behaviors take place and in which research takes place. This Linguistic/Intensive view emphasizes the subjective (cf. Bachman, 1990, pp. 37-8, 324) or variable (see section 2.2) nature of knowledge, and often uses a conceptual, sociolinguistic, or functional analysis of language. These Linguistic/Intensive approaches therefore attempt to account more fully for the complexities of language use indicated by these linguistic theories, often using transcribed spoken language data as one of several data types. Within this Linguistic/Intensive approach, reliability and quantification are often not emphasized in order to emphasize construct validity, authenticity, individuality, or a view of language as a functional,

open system (see especially Lantolf & Frawley, 1988). A Linguistic/Intensive approach to research often uses complex time-consuming analyses and longitudinal designs, with a large number of measurement instances. These choices permit only small samples or individual cases to be studied. The normal purpose of a case study is to understand the case rather than to make generalizable conclusions about causality. As a result, the use of very small samples usually precludes the possibility of rigorous investigations of causality. The studies of language development in Klein and Perdue (1988) and Huebner (1979, 1983a, 1983b), and the instructional case studies of Pienemann (1984) and Schumann (1978) are examples of this second type of approach. The limited internal and external validity of this approach has been discussed, for example, in critiques of Pienemann (1984) by Ellis (1990a, pp. 155, 158) and by Krashen (1992, 1993).

The Psychometric/Extensive approach has dominated the research designs chosen for L2 instructional studies, and these designs have, consequently, provided relatively few studies with mid-term and long-term unmonitored post-tests. In this dissertation and in Mellow, Reeder and Forster (in press), it is argued that these two types of approaches can be integrated, achieving generalizations, although the generalizations are analytic (to theory) rather than statistical (to populations) (Yin, 1989, pp. 21, 38-40), and balancing linguistic and psycholinguistic concerns for spontaneous longitudinal data with concerns for reliable measurement and internal validity (cf. Bachman, 1990, pp. 296-334; Kazdin, 1982).

3.3 The Time-Series Design

A design that balances intensive and extensive approaches is the single-subject time-series design. Time-series (T-S) designs may practically combine features of the two previously-noted approaches, resulting in increased construct validity and reasonable internal validity. Mellow, Reeder and Forster (in press) provide a detailed discussion of T-S research designs and their features, as well as a discussion of their use for SLA research. A T-S design was utilized in this study. In this section, a number of properties and features of the T-S design are discussed. The

discussion in the section focuses on the rationale for the specific features and methodological strategies adopted in this dissertation study. After this general discussion of these features, the specific details of the method are described in Chapter 4.

A T-S design involves the use of many parallel, reliable measurement instances for each participant in a study, including a number of instances *before* the intervention. These repeated pretests form a baseline period and constitute the most important property of T-S designs because they allow researchers to determine the typical behavior expected for a participant, providing a comparison for postintervention behaviors.¹³ If the preintervention baseline behaviors are consistent and stable, then researchers may be relatively confident in attributing causality regarding postintervention changes in behavior to the treatment. In this way, internal validity is increased because participants act as their own controls. Given the non-equivalence of groups that is problematic for most quasi-experimental studies, discussed in sections 3.2 and 4.1.1, the use of participants as their own controls is one of the strengths of this design: the baseline comparison is an exactly equivalent group. In addition, T-S designs typically include a large number of measurement instances after the intervention. These multiple posttests permit an examination of longitudinal effects (cf. Cook & Campbell, 1979, pp. 225-28), described here as the developmental course of SLA (section 2.3). In addition, because a small sample is studied, this facilitates the use of several types of measurement instruments, including recordings and analyses of spoken data, allowing an examination of the unmonitored language use.

An important aspect of the time-series design that was considered in this study was the time between measurements (cf. Borg & Gall, 1989, p. 716; McCleary & Hay, 1980, p. 21). In principle, it would have been desirable to obtain language samples as often as possible to increase the number of observation points (to increase reliability and the possibility of using statistical tests)

¹³ The term *baseline* refers to any period in which treatment is not being provided. While the term is often used to refer exclusively to the pre-test period, it may also refer to a period after an intervention has been withdrawn, that is, a return to baseline. A variety of design options are possible, including AB, ABA, and ABAB, with *A* denoting a baseline phase and *B* denoting a treatment phase (cf. Mellow, Reeder & Forster, in press). Because the postintervention phase in this study included sessions of instruction for three learners, the postintervention phase will not be described as a baseline phase (cf. sections 4.1.1 and 4.1.2).

and to record changes that may have been rapid or temporary. However, the overtesting of individuals might have resulted in boredom, practice effects, or excessive demands on the learners' time (resulting in the withdrawal or attrition of the volunteer participants) and therefore testing might have threatened the internal experimental validity (e.g., Kazdin, 1982, pp. 78-9; Kratochwill, 1978, p. 14). In the study reported here, two weeks was chosen as the period between measurements in order to balance these two concerns. An attempt was made to keep the time between observations relatively constant throughout the study in order to make the results consistent and reliable. Although consistent time intervals were desired, the practical concerns that resulted from using volunteers compromised this ideal (cf. sections 4.1.1 and 4.1.2).

A second important aspect of the time-series design that was considered in this study was the length of the preintervention baseline period. Ideally, this period should include as many observations as possible. Glass (1980) indicates that about 15 baseline measurements would be desirable (cf. Cook & Campbell, 1979, pp. 228-30). However, that number of observations was impractical in this study. In addition, the desired length of the baseline may also ideally be determined by the stability of the results, that is, the baseline should continue until the scores are relatively consistent. Without stability in the baseline, it is difficult to establish clear effects of the intervention. However, given the variable nature of interlanguage use and the continued development that is expected of learners living and studying in the SL environment, stable baseline data may be difficult to obtain (for discussions of trends and variability, see Kazdin, 1982, pp. 106-109, 263-269; Kratochwill, 1978, pp. 81-3). In addition, because of the time-consuming nature of data analysis, it would have been difficult to constantly analyze the data during the study to determine if stability had been achieved. Therefore, in this study, it was decided that the baseline would last for approximately half of the period of study (cf. section 4.1.2). A discussion of strategies that were used to increase the stability or reliability of the scores is provided in section 3.3.3.2.

3.3.1 *Time-series Designs: Internal Validity*

Internal validity, the valid attribution of causality, is crucial for any experimental or quasi-experimental study. As noted above, one of the strengths of the time-series design is that participants act as their own controls, providing a baseline that is an equivalent control group. Three additional design features were used in this study to further increase internal validity: replications, multiple-baselines, and a control group (for additional options, see Mellow, Reeder, & Forster, in press).

The first design feature used in this study to increase internal validity was the inclusion of four learners who received treatment. Each learner was a replication of the single-case design (e.g., Kazdin, 1982, pp. 284-287; Kratochwill, 1978, pp. 22-23; Tawney & Gast, 1984, pp. 94-109). This number of replications is more extensive than previous instructional studies that have employed case study approaches (e.g., one individual: Schumann, 1978 and Kadia, 1988; two individuals: Pienemann, 1984).

The use of replications is especially valuable if a second feature, multiple-baselines, are used (e.g., Borg & Gall, 1989, pp. 720-23; Cook & Campbell, 1979, 222-25; Kazdin, 1982, pp. 126-229; Kennedy, 1988; Kratochwill, 1978, esp. 71-3; Mellow, Reeder & Forster, in press; Tawney & Gast, 1984, pp. 226-340). In a multiple baseline design (across learners rather than across behaviors), each learner randomly receives the experimental treatment session (or the onset of treatment) at a different point in time (i.e., in terms of chronological time or in terms of the number of baseline observations). These designs reduce the threats of history and maturation: external (environmental) events (such as a lesson in another class) or internal development may possibly coincide with one introduction of a treatment, but any consistent treatment effects in the learners in a multiple baseline design are most likely due to the intervention. Thus, a multiple baseline study increases confidence in the causal relationship between treatment and changes in language development. In this study, a multiple-baseline was used, because the learners received instructions on different days, although the length and timing of the multiple baselines were not

entirely predetermined systematically, as is done in many multiple baseline designs (cf. sections 4.1.1 and 4.2.2.2, and Mellow, Reeder & Forster, in press).

The third design feature used in this study to increase internal validity was the inclusion of a non-equivalent no-treatment control group. If the treatment and control groups experienced a similar set of external events during the study (i.e., their *history* was the same) and if their pre-intervention baseline behaviors were similar, suggesting that they will develop in similar ways, then the scores of the control learners can be used to assess the effects of the treatment. However, this design option was somewhat problematic because the treatment and comparison groups differed from each other, an issue discussed in sections 3.2, 4.1.1 and 5.1.

Internal validity may be particularly threatened in T-S designs (or any experimental design) because dependent variables may exhibit a systematic change or trend during the course of the experiment: Continuous development is expected for learners who are acquiring the language spoken in the society around them or who, in their language classes, are learning the language and are becoming better test-takers (i.e., history threatens internal validity, cf. section 2.3). If the preintervention baseline phase indicates a gradual increase in performance, then the postintervention phase may exhibit higher, more target-like scores due to this continuing developmental trend rather than due to the treatment. In fact, this trend is problematic in all studies of instructed SLA and appears to be the reason that long-term effects of instruction were not found in Harley (1989a): Learners in the control group continued improving until they caught up with the treatment group.

As indicated in section 2.3, the use of a control group is perhaps the primary method of assessing long-term effects of instruction if a developmental trend is evident in the preintervention baseline. Two additional strategies were used in this study to diffuse the problem of trend in the preintervention baseline (for further discussion, see Mellow, Reeder & Forster, in press). First, an attempt was made to ensure that the dependent variables were as precise and detailed as possible (cf. Glass, 1980). By doing so, changes in behaviors were likely to be larger and more dramatic because the changes were not being mediated or diluted by other behaviors. Therefore, based

upon the discussions in Chapter 2, in addition to analyzing the L2 use of English articles in general, this study investigated the potential effects of article instruction on the use of *the* suppliance in specific syntactic contexts, including subject and non-subject positions, and for differing degrees of referential distance.

A second strategy that was used to diffuse the problem of trend in the preintervention baseline was to accurately compute the slopes of regression lines that characterize the trend in the group scores during the preintervention and postintervention phases (cf. section 4.2.2.2). The use of these lines made it possible to precisely assess changes in trend and level. Although a slope line does provide valuable information about development, it may not necessarily be the most accurate representation of development because SLA is not linear and instead is characterized by restructuring, backsliding, U-shaped development, and variability. As a result, an analysis of individual line graphs is provided in section 5.3 to complement the information provided by the regression lines.

Because I was both the researcher and the instructor of the intervention in this study, this resulted in an additional threat to the internal validity of the study, especially because my presence may have suggested to the learners that the postintervention sampling sessions were an opportunity to continue practicing article use. Although this situation is not unusual in single-subject research (cf. Ishiyama, 1991, regarding the scientist-practitioner paradigm), an attempt was made to carefully document all instruction and to probe for attention to articles during a final interview in order to assess the impact of this threat.

3.3.2 Time-series Designs: External Validity

An important concern for the interpretation of this study is the issue of external validity, the degree to which the results can be generalized beyond the conditions of the experiment. Although external validity is clearly a weakness in this T-S design or any other small sample design, and although this prohibits generalizing the results to larger populations, the T-S design employed in

this study may be equal or superior to control group designs with respect to certain aspects of external validity (cf. Mellow, Reeder & Forster, in press).

First, and perhaps most significant for this instructional study, the T-S design appeared to have controlled for a very serious threat (or set of threats) to external validity. This threat, which I refer to as the Hawthorne Effect, may not be controlled in other designs. Also known as reactive experimental arrangements, reactive assessment, and pretest sensitization (e.g., Kazdin, 1982) or the effect of the research environment (Seliger & Shohamy, 1989), the Hawthorne Effect arises when learners know they are being tested and therefore behave differently (for a detailed discussion, see Borg & Gall, 1989). This problem is especially acute if learners are also aware of the exact behavior being tested. If this occurs in SLA studies, such studies clearly cannot assess the unmonitored use of language. Participants may become aware of the behavior being studied when either the tests obviously indicate the language element under investigation or when the treatment is somewhat unusual, or both, thereby alerting participants that the researchers are investigating the element or behavior that is being taught. When these threats are present, researchers cannot be certain that instructional effects would occur outside of the experimental context (cf. Krashen, 1993). In addition, if the tests are designed in such a way that the variables being investigated are obscured, but the focus of the treatment is obvious, then this would constitute a serious threat to internal validity in a control group design because the experimental groups would know the aspect being studied and the control group would not, behaving differently during measurement sessions.

As indicated by the focus on unmonitored language use in the research hypotheses in section 2.5, one of the primary concerns that guided all methodological choices was a desire to avoid the Hawthorne Effect. Therefore, efforts were made to prevent learners from discovering the aspect of language being studied so that, when producing their language samples, the learners did not focus on and attend primarily to article use. These strategies included: (a) the use of self-selection as the method of determining both the composition of the treatment and control groups and the number of instructional sessions received by each instructed learner (section 4.1.1; cf.

strategy (d) below); (b) the use of long baseline periods, with multiple sampling sessions in each, before and after the instructional periods (section 4.1.2); (c) the use of measurement instruments that did not draw learners attention to article use: spoken and written retellings, as well as cloze tasks with many distractors (sections 4.1.3 and 4.1.4); (d) the provision of many other types of non-compulsory instructional activities during the period of the study, so that there would be no special emphasis on the article instruction (section 4.1.5); and (e) the justification of the provision of the experimental treatment as a response to the requests of the learners (section 4.1.5). Because the Hawthorne Effect may be the single most serious threat to studies of unmonitored language use, its reduction through these strategies took precedence over other design requirements, such as the need for strict randomization, that might have permitted specific statistical analysis of multiple baselines (cf. Kennedy, 1988; Mellow, Reeder & Forster, in press).

Although the results of this study cannot be generalized across learners or settings to a larger population, large sample nonequivalent group designs also may not be generalizable because samples cannot be selected randomly from a defined population, and instead intact groups are often used (e.g., Runkel, 1990, p. 20; Woods, Fletcher & Hughes, 1986, pp. 48-57). Consequently, increased external validity for all quasi-experimental studies is normally achieved by careful replications. Therefore, in this study an attempt has been made to carefully describe baseline and treatment conditions, characteristics of learners, and measurement procedures (Chapter 4), so that a basis will exist for determining the degree of generalizability of results from this study to other studies. In addition, by carefully manipulating the variables under investigation and addressing key theoretical and methodological issues, the value of this study may primarily lie in generalizing to SLA theory (analytic generalizations) rather than to defined populations (statistical generalizations based on the normal distribution of scores) (cf. Yin, 1989, pp. 21, 38-40; cf. Larsen-Freeman & Long, 1991, p. 14; Runkel, 1990). In other words, the results of this study may be used to construct more precise theories, especially transition theories that consider the many variables affecting language learning. This appears to be the approach to external validity adopted by Cummins (1987, p. 192) in his evaluation of immersion research:

In this paper I will . . . summarize the findings of research on immersion. Then the theoretical principles that emerge from the immersion research together with other research on bilingual education will be outlined. It is crucial for language planning purposes to make these theoretical principles explicit because it is only on the basis of these principles that predictions can be made about academic outcomes under different social and educational conditions.

In sum, although external validity is clearly a concern for this study, the use of this design appears to be valuable because of the possible control over the Hawthorne Effect, the concern for variation in individual cases (cf. sections 2.2, 3.2, and 5.3), the feasibility and practicality of replications, and the possibility of generalizations to SLA theory. In addition, by providing valuable information that is difficult to obtain using other designs, this time series design can be used to inform and complement other types of quasi-experimental and case study research.

3.3.3 Time-series Designs: Measurement Issues

In section 3.2, it was suggested that the small number of instructional studies employing measures of unmonitored, spoken language use was due to the dominance of a Psychometric/Extensive approach to instructional studies. The use of a psychometric approach in previous studies led to an emphasis on the reliability of measurement instruments for assessing development, ensuring that different versions of a measurement instrument were parallel and reliable so that claims about changes in behavior could be made. As a result, written metalinguistic tasks, such as judgements and corrections, were often used because parallel versions can be created, administered, scored, and analyzed in a relatively efficient and reliable manner (cf. Bachman, 1990, pp. 57, 298, 314). However, it is possible that the validity of measurement instruments was limited in order to maintain reliability. Given this previous emphasis in instructional studies, this section considers the measurement instruments that were chosen for this study, indicating the attempts that were made to find a balance between validity and reliability.

3.3.3.1 *Establishing the Validity of Instruments*

A test or measurement instrument is valid if it is meaningful, appropriate, useful, and authentic (Bachman, 1990, pp. 25, 236-237, 296-303). The difficulties in validation lie in defining relevant terms or constructs, such as *unmonitored language use*, and then providing theoretical or logical evidence (content validation) and empirical evidence (criterion validation) that the instrument precisely measures these defining characteristics (cf. Bachman, 1990, pp. 41-44, 236-271; Cumming & Mellow, 1996; Standards for Educational Research, 1985, p. 9). There are no clear criteria for determining when construct validity has been achieved for any specific instrument. The approach to validation used here was to establish content validity, that is, to look initially to educational goals (Chapter 1), linguistic theory (including the desired analyses; Chapter 2), and previous instructional studies (section 3.1) for an appropriate theory of language development to inform the instruments (cf. Bachman, 1990, pp. 244-247, 302-303, 315-333).

More specifically, the theory adopted here has defined language development in terms of interlanguage form-function mappings and interacting cognitive and linguistic principles. The hypotheses derived from these principles required: (a) a narrative monologue to elicit discourse that included referent tracking, subject and non-subject positions, and variations in referential distance; and (b) a range of tasks that required, allowed, and prevented learners from attending to form and drawing upon consciously learned rules. Consequently, the study adopted story retellings (specifically fable retelling), which have been used in previous linguistic and SLA studies (e.g., Chafe, 1980; Fakhri, 1989; Klein & Perdue, 1988; and Swain, 1985), and also adopted specific analyses of article use that drew upon the analyses in Chaudron and Parker (1990), Givon (1983), and Huebner (1979, 1983a, 1983b). Section 4.1.4 provides a more detailed discussion of the properties of the measurement instruments.

3.3.3.2 *Establishing the Reliability of Instruments*

Because the present study of instructed language acquisition requires more than one session of measurement, the reliability of the different versions of the measurement instrument was

crucial. Without reliability, it would not be possible to determine if changes (or the absence of changes) reflect the ability in question (the true score) or are due to other factors which affect the reliability of the instrument (the error score; e.g., increased familiarity with the test, differences in putatively parallel tests, or unsystematic variability in the learner) (e.g., Bachman, 1990, pp. 166-69). To reveal very large effects, a crude, somewhat unreliable instrument may be sufficient (Borg & Gall, 1989, pp. 257). However, because the effects of instruction have been so difficult to determine (Chapter 1 and section 3.1), the effects found in instructional studies may be quite subtle, and therefore the psychometric property of reliability is very important. In addition, because the T-S design requires a relatively large number of parallel versions of assessment instruments, the reliability of the measurement instruments took on even greater importance. Finally, because an investigation of developmental course, as investigated in this study, required, at each sampling session, the use of multiple tasks that were equivalent except for varying degrees of attention to form, consistency of measurement was extremely crucial.

Two types of reliability were sought in this study: (a) reliability in the transcription and coding of the data, and (b) reliability or equivalence in the parallel forms of instruments. The first concern, establishing reliability of transcription and coding, is especially important given the nature of functionalist interlanguage analysis. The judgments involved in interpreting interlanguage article use are very complex (cf. section 2.1.2, and Cumming & Mellow, 1996, pp. 78-80). As discussed in section 2.1.2, as many as 21 functional categories of target contexts have been used to analyze L2 article use (Thomas, 1989), and because *a*, *the* or \emptyset might appear in each context, each of the 2265 nouns in Thomas' corpus was categorized in one of 63 possible ways. Given the potential ambiguities in such classification systems, especially in terms of understanding the speaker's intentions, reliability may be highly problematic. Indeed, as a result of the difficulty in determining these distinctions, Brown (1973, p. 350), in his influential work on the acquisition of grammatical morphemes, remarked that distinguishing these semantic contexts "proved to be not possible." In order to ensure reliability, Cumming and Mellow (1996) simplified the decision making process of the raters in that study of article use. The decision making process of the raters

in the study reported here was also fairly simple and involved very little inference regarding speakers' intentions (cf. section 4.2.1). However, to increase the reliability of coding, two coders and coder training were used.

The use of multiple coders is often advocated as an important method for increasing reliability (e.g., Bachman, 1990, pp. 180-1; Borg & Gall, 1989, pp. 525-8; Kazdin, 1982, pp. 48-75; Seliger & Shohamy, 1989, pp. 185-7). Given the complexity of interpretation of article use, it is surprising that very few previous analyses of interlanguage article use reported multiple coders, coder training, inter-coder reliability, or intra-coder reliability (e.g., Huebner, 1983a, pp. 38-40, 1983b, pp. 52-53; Klein & Perdue, 1988, pp. 11-22; Master, 1987, pp. 29-42; Parrish, 1987, pp. 363-365; Pfaff, 1987a, pp. 88-89; Tarone & Parrish, 1988, p. 28). Thomas (1989, p. 345) reported inter-coder reliability for the identification of the items in the text, but not for the probable choice of articles by native speakers. In Cumming and Mellow's (1996) study of L2 article usage in compositions, two raters categorized each of 2085 items according to four categories, and achieved 93% agreement on 100% of the corpus. The training, conventions, and agreement of the two coders in the present thesis are described in detail in section 4.2.1.

The second type of reliability that was sought in the present study involved the determination of the equivalence of parallel forms of the instruments. The equivalence of the parallel forms of a test is often considered to be one of three types of test reliability, with test-retest and internal consistency as the other two types (e.g., Bachman, 1990, pp. 172-187; Borg & Gall, 1989, pp. 257-262; cf. Seliger & Shohamy, 1989, pp. 185-187). For the longitudinal research reported here, parallel forms were required rather than retesting using the same test because a practice effect and boredom would have occurred. An extensive discussion of parallel test reliability is provided in Ghiselli, Campbell, and Zedeck (1981, pp. 192, 199-205, 244-245, 249-251, 259-260).

Establishing reliability for tests of dynamic, changing traits, such as language ability, is both very important and very difficult (Bachman, 1990, pp. 181-3; Ghiselli, Campbell & Zedeck, 1981, pp. 244-5, 251, 259). Because change in the trait may affect the scores, and hence make

reliability co-efficients impossible for parallel tests, Ghiselli, Campbell, and Zedeck (1981, p. 260) suggest that internal consistency tests (e.g., split-half correlation co-efficients) should be used to establish reliability rather than equivalent forms co-efficients. However, this does not resolve the problem of equivalent tests, and only ensures that each test is reliable internally. Furthermore, testing internal consistency is untenable for the types of instruments used in this study, because the narratives used to elicit the tracking of referents form a single unit and do not have equal subparts.

In order to improve the equivalence of parallel versions of instruments, four strategies were used (for further strategies and discussion, see Cook & Campbell, 1979, pp. 43-4; Ghiselli, Campbell, & Zedeck, 1981, pp. 261-2). First, based on theoretical and empirical insights, the measurement instruments were designed in order to control and keep constant as many variables as possible. The references listed in section 2.2 provide discussions and inventories of discoursal and situational aspects of the context of language use which may be controlled, including time for planning, the relationships of the interlocutors, the topic, the conventions of turn-taking and rhetorical organization (cf. Klein & Perdue, 1988, pp. 11-22; Pica, 1983b; Tomlin, 1990, pp. 165-6). For a time-series study of article use for referent tracking, both discoursal and situational factors must be controlled. With respect to the linguistic demands of each task, the narratives were kept comparable in terms of: (a) the approximate number of NPs that were non-generic, countable and singular; (b) the overall length, complexity, and cultural biases of the content; and (c) the number of characters or referents. With respect to situational factors, the test conditions were kept highly comparable (cf. test method facets, Bachman, 1990, pp. 111-59): (a) all measurement sessions (except one) occurred in the same physical setting, a language laboratory; (b) the same interlocutor was always present; and (c) for the fable re-tellings, the amount of background information was kept constant, including the number of pre-listenings, the opportunities to ask questions about vocabulary and content, and the support materials, summary graphics. Section 4.2.3 provides a detailed discussion of parallel procedures for data elicitation.

The second strategy for improving the reliability of parallel versions of the test instruments was the use of counterbalancing (Borg & Gall, 1989, p. 711; Bachman, 1990, p. 183).

Counterbalancing, systematically alternating the order in which the test versions are given to learners, should minimize the effects of specific properties of any one version of an instrument. In particular, the use of different versions of the instrument immediately after intervention is important, so that consistent changes can be attributed to the intervention rather than to properties of that version of the instrument. The counterbalancing of the fable versions is described in section 4.1.4.

The third strategy for improving the reliability of the scores derived from parallel tests was the use of instruments that had a relatively large number of contexts for *the* suppliance: with a large number of contexts, inconsistencies may have had a less significant effect on the scores (e.g., Cook & Campbell, 1979, p. 43; Cumming & Mellow, 1996; Ghiselli, Campbell & Zedeck, 1981, p. 261; Seliger & Shohamy, 1989, p. 187; cf. Borg & Gall, 1989, pp. 267-9). For example, Tarone and Parrish (1987, p. 31) included data in which a learner had only two contexts in which a form was required: a single inconsistency changed a learner's score by 50%. The reliability of these scores was increased, however, when Tarone and Parrish averaged the results of the learners to determine a group mean (hence reducing individual variability).

Fable retellings were adopted in this study to maximize the number of contexts for *the* suppliance. Although these tasks resulted in 33% of the scores being based upon 10 or more contexts for a dependent variable, because four of the dependent variables required fairly specific linguistic contexts, 8% of the scores that were reported were based on only three contexts (cf. section 4.2.2.1). Therefore, this strategy for increasing reliability was not entirely successful in this study.

The fourth strategy that was used to increase the reliability of the scores derived from the parallel tests was to replicate the study with additional learners and then reduce variation in the scores by using group mean phase scores and plotting regression lines using group scores (Cook & Campbell, 1979, p. 43; Walter Boldt, personal communication) (cf. section 4.2.2.2).

Overall, reliability, like validity, can never be firmly established in behavioral research but must be based upon the accumulation of logical and empirical research (Bachman, 1990, p. 161).

Therefore, these four strategies were used to improve the reliability of the measurement instruments and to increase confidence in the results. However, all results in behavioral research, especially those assessing the nature of development, must be interpreted with caution.

3.3.3.3 *Establishing Appropriate Measurement Scales*

In addition to concerns about validity and reliability, the choice of instruments and analyses was influenced by an additional psychometric principle: the use of appropriate measurement scales (e.g., Ghiselli, Campbell & Zedeck, 1981, pp. 9-15, 31-41, 57; Bachman, 1990, p. 24-27, 30-40, 296-334). Certain types of measurement scales are preferred for psychometric reasons. The most desirable types of scales, ratio and interval scales, are those which allow the variable (i.e., the operationalized characteristic or property of the individual) to be measured with equal units along the scale (Ghiselli, Campbell & Zedeck, 1981, p. 31-2, 34). With the use of this type of scale, a precise comparison can be made of individuals who achieve different scores on the instrument. This type of scale is used in this study when *the* suppliance is expressed as a percentage of the total number of obligatory contexts for that form.

However, language use and development cannot necessarily be represented by units that are equal in this way (e.g., Bachman, 1990, pp. 26-30). Language use may be too complex to be effectively represented by one interval variable or even a set of interval variables, and therefore the use of psychometric principles to evaluate language development and proficiency has recently been critically evaluated (Lantolf & Frawley, 1988, pp. 186 - 90). In their discussion of the measurement of oral proficiency, Lantolf and Frawley (1988, p.185), citing Bissert (1979) and Lewontin, Rose, & Kamin (1984), consider the problem of the *psychometric posture*, suggesting that language proficiency research has been hindered by the domination of psychometric principles, that researchers have lost "sight of the object of scientific inquiry in order to preserve the integrity of the tools they use to measure the object", and that even though it may be possible to devise instruments with metric scales, that does not mean that the ability being measured by the instrument is really metric.

The limitations of metric scales for measuring language development can be illustrated with examples of the issues discussed in Chapter 2. First, because language use is variable, changing systematically in response to a multitude of situational or task related variables and varying in unsystematic ways as well (e.g., Ellis, 1989; Tarone, 1988, pp. 111-114), this inherent variability and the difficulty in isolating all of the factors which affect variability suggest that language ability may not be metric. Second, the studies reviewed suggested that language development is not merely a linear accumulation of rules or elements in which development occurs in equal increments. When measured along an interval scale, language development proceeds in bursts and jumps, often backsliding before further progress can be made. This is most obvious in U-shaped development or restructuring, exemplified in the overgeneralizing stage of *the-flooding* that is often found in L2 article development (stage 3 in Table 2.2). Similarly, stages of development are not equal because early development is often rapid, but then development may slow and ceiling effects may occur because the achievement of 100% of target-like use may be confounded by a variety of factors, including interference with other linguistic subsystems, motivation, and social distance (cf. Principle 6 in section 2.5).

For these reasons, the use of interval scale scores, such as percentage of target like use, do not always ensure that the units are equal. The difference between 20% and 30% accuracy may not be comparable to the difference between 80% and 90% accuracy. In addition, a change from 20% to 30% may not represent a comparable change in different learners.

Because of these issues, three different types of analyses are used in this study (cf. section 4.2.2.2). First, group mean phase scores and inferential parametric statistical tests are used, utilizing interval scales. However, the difficulties in using such measures and scales is discussed in sections 5.2.1. In addition, the slope values of regression lines are used for the analyses in section 5.2.1 and 5.2.2. Although the slope values use an interval scale, because they measure developmental trend rather than providing a score on a scale of 0 to 100, they are better able to reflect the nature of L2 development. Finally, line graphs of individual scores are used in section 5.2.3. Once again, the individual session scores use an interval scale, but the purpose of the

analyses is to examine the nature of developmental course, considering, for example, the possibility of U-shaped development (cf. section 2.3). In sum, the importance of measurement scales for assessing development is examined throughout this study.

Chapter 4

Method

The three previous chapters have provided a review of pedagogical, linguistic, cognitive, and methodological issues related to conducting instructional studies of second language acquisition. These issues have motivated the research hypotheses and design of this study. This chapter describes in detail the method and procedures used in the study, including descriptions of the learners in the treatment and control groups, the design, the data collection procedure, the measurement instruments, and the instruction. Chapter 4 also describes the nature of the data analysis, including descriptions of the transcription and coding, definitions of the dependent variables, and explanations of the statistical analyses that were chosen to summarize the group data in terms of central tendency and trend.

4.1 The Method of the Study

The study used a multiple baseline, non-equivalent control group time-series design to investigate the effects of form-focused grammar instruction on the L2 development of article use. In this longitudinal multiple baseline design, eight adult Japanese learners of English provided language samples approximately every 2 weeks throughout a 6 month period, each providing between 6 and 11 separate sets of 3 language samples. During this time, the learners were living in Canada, with English-speaking room-mates, and were taking theme-based university preparation ESL courses at a Canadian university. The learners were recorded individually, and during each recording session (except for slight variations in the first two sessions) provided three samples of language: a spoken retelling of a fable (abbreviated Sp), a written summary of the same fable (abbreviated Wr), and a written completion of a cloze passage of that fable (requiring six articles and seven to nine distractor items; abbreviated Cl). These retellings were based upon the learners' memories after listening to the fable twice, with minimal content-oriented support from a time sequence graphic that I provided. These three tasks were chosen in order to vary the amount of

attention to specific forms (cf. Tarone & Parrish, 1988). Each learner therefore provided as many as 31 samples (e.g., 11 spoken fable retellings, 10 written fable retellings, 10 cloze tests). The language samples were transcribed or transferred to computer documents, and then coded for several specific aspects of article use.

After a 3 month baseline period, four of the eight learners were given instruction in article usage, focusing on the use of articles to track referents in narrative discourse. Learners received one, two, or three sessions of instruction. Through the use of extensive baseline data (i.e. multiple pretests: four, five, or six) the learners acted as their own controls and each learner's data set is considered to be a replication of a single-subject design. The data from these learners (the treatment group, abbreviated TrG) is also discussed in comparison to data supplied by the four uninstructed learners (the control group, abbreviated ConG).

4.1.1 Learners

The non-native speaker learners were eight volunteers (four male, four female) who are adult Japanese learners of English and who, at the beginning of the study, were between 19 and 22 years of age. These learners were Japanese university students who were in Canada for 8 months as part of a university exchange. All eight learners had very similar backgrounds, having lived and been educated in Japan, and having chosen to live and study in Canada for 8 months. During the first 4 month semester, the learners took three courses that might best be described as theme-based ESL instruction (Brinton, Snow, & Wesche, 1989). The courses were taught by language instructors rather than content instructors. The themes were related to topics that might be studied in university Arts courses. The classroom materials and activities had been modified to foster language learning and therefore were similar but not equivalent to those typical of university courses for native English speakers. The courses did not typically include form-focused (grammatical or rhetorical) instruction. During the second 4 month semester, the students took between three and five courses, depending on their TOEFL score and their wishes. These second term courses included a combination of any one university credit course, two *core courses*

(sheltered content courses taught by content teachers, i.e., university faculty in specific academic areas), and three theme-based courses taught by ESL instructors.

Because informed consent was obtained from these learners (following procedures outlined by the university's ethical review committee), the learners were clearly aware that they were volunteers in a research project and that they could quit the project at any time and could participate to the degree that they wished (see Appendix A). As a result, the volunteers did not all participate fully in the project. In fact, the project began with 18 volunteers and had regular involvement from 10 volunteers for the first 3 months. In addition, the manner of group selection was not random: The eight learners described in the study were simply offered as many as three sessions of instruction, and therefore the groups were self-selected. One learner chose to participate in one session, two learners chose to participate in two sessions, and one learner chose to participate in three sessions. Therefore, these four learners became the experimental group. As a result of busy academic schedules and other reasons, four learners chose not to receive instruction in article use, thereby becoming the control group. Although the four uninstructed learners probably would have found time for the instruction in article use if I had made a clear or emphatic request, such a request was not made in order to avoid the Hawthorne Effect: Every effort was made to prevent the learners from knowing that article use was being investigated and consequently focusing on articles during the sampling sessions.

Although the gender distribution remained equal in both groups, this method of group selection resulted in non-equivalent groups. In addition to concerns about non-equivalence in the learners' levels of motivation and interest in grammatical accuracy, the learners were different in several more easily quantifiable respects. The best TOEFL score of the learners in each group varied somewhat, with instructed learners ranging from 537 to 583 and averaging 555, and uninstructed learners ranging from 483 to 510 and averaging 504. Oral Proficiency Interviews (OPI) and Test of Written English (TWE) standardized test scores, from the beginning and end of the 8 month exchange also exhibited differences. For the instructed learners, the initial OPI scores were 1, 1.9, 3, and 3.9, changing to 2.9, 2, 3.9, and 3 at the end of the study. For the

uninstructed learners, the initial OPI scores were 0.9, 1, 1.9, and 2, changing to 1.9, 2, 3, and 2 at the end of the study. For the instructed learners, the initial TWE scores were 2, 2, 3, and 3, changing to 3, 4, 2, and 4 at the end of the study. For the uninstructed learners, the initial TWE scores were 1, 2, 2, and 3, changing to 2, 3, 3, and 3 at the end of the study. Thus, in general, the instructed group performed at higher levels on standardized tests than did the uninstructed group, both at the beginning and end of the 8 month period of the study.

In addition, the two instructed learners who had initial OPI scores of 3 and 3.9 had lived for short periods of time in the United States, totaling 10 weeks during high school, and 16 months during elementary school and junior high school, respectively. Finally, whereas the instructed learners provided either 28 or 31 samples of data, the uninstructed group did not provide data as regularly, providing 17, 18, 18, and 24 samples (the distribution of these sessions is illustrated in Appendix B). Therefore, although all eight learners were similar in many respects, the self-selection procedure in this study resulted in two non-equivalent groups. This non-equivalence is explored further in section 5.1.3. However, the important implication of this non-equivalence is that the strongest claims for causality in this study can be made not with a between-group comparison, but instead when the four instructed learners act as their own controls, comparing their preintervention behavior to their postintervention behavior (i.e., Hypotheses 7a and 8a).

Six of the students, including all four instructed learners, participated in a final interview during their last week in Canada. I took notes and tape-recorded their responses. In that interview, which lasted between 45 and 75 minutes, the learners provided the following types of information: 1) a description of their personal background (including previous language learning experiences), 2) an assessment of their own personality and language learning style, 3) a detailed description of their activities in and for their courses at the university, 4) a detailed description of their English speaking activities outside of class during their stay in Canada, and 5) a description of their impressions of the language elicitation procedures, including their predictions as to the topic of this research.

In addition to indicating the relative homogeneity of those six learners, the interviews revealed two findings that are important for assessing the internal and external validity of the study. First, in the interviews, the learners said that they had received no other instruction regarding article use during the 8 month period of the study and that, in their classes and their assignments, grammar and error correction had not been emphasized. Second, none of the learners reported that they suspected that articles were the topic of the study and none of the learners reported that they focused significantly or primarily on articles during the sampling sessions. When asked to predict the topic of the study and whether there had been any specific aspects of language that they had focused on during the sampling sessions, the four instructed learners all mentioned articles, but only as part of a list of other items, including word choice, speaking, pronunciation (including tempo of speech and absence of pauses), accent, organization, vocabulary, grammar, tense (especially past tense), prepositions, and number marking. Neither of the two uninstructed learners who were interviewed mentioned articles as a predicted topic of the study or as a focus during sampling sessions, but they did mention a number of the categories listed above, and one learner reported that she had focused on gender and possessives as well. In sum, the instructed learners did not appear to have changed their behavior in the sampling sessions because they thought their article use was a distinct topic that was being studied.

4.1.2 Design

As discussed in Chapter 3, the instructional study in this dissertation is a quasi-experimental time-series design. In order to answer the research questions outlined above, this design permitted an intensive analysis of a fairly extensive data set for each learner. The design increased internal validity because the learners acted as their own controls and were compared to a non-equivalent control group. The learners were encouraged to provide a data sample every two weeks (cf. section 3.3). However, as indicated in Appendix B, the time between sampling sessions varied somewhat, and, especially for the control group, the time between sampling sessions was often longer than two weeks during the second half of the study.

The four replications of the instructed learners resulted in a multiple baseline design (cf. section 3.3.1), with the learners receiving instruction on the following days of the study, with day 1 being the first day of a sampling session for any learner (October 27th, 1992): 93; 95 and 109; 97 and 118; and 86, 92 and 108. Therefore, each of these eight instructional sessions occurred on a different calendrical date during a one month period. Although I attempted to vary the dates on which instruction was provided, variations also resulted from the different times that the learners were available to be recorded and to receive instruction (i.e., because of class and holiday conflicts). Thus, the schedule of treatment was somewhat uneven, in response to the various weekly schedules of the learners, rather than being random in a predetermined systematic fashion.

The data in this study were divided into two phases, the preintervention (PreI) phase and the postintervention (PostI) phase (cf. section 3.3, especially Footnote 13). The PreI phase was the period from the first sampling session until the last sampling sessions before instruction. The PostI phase was the period from the first sampling session after instruction until the last sampling session of the study. When all of the sampling sessions for the learners in the treatment group were combined, the PreI phase for that group was days 1 to 90, and the PostI phase was days 92 to 173, with the day numbers relative to the first day of the study. A comparable phase division was determined for the control group, with the PreI phase being days 1 to 78, and the PostI phase being days 85 to 173.

The PostI phase was defined as beginning with the first sampling session after instruction. For two of the instructed learners, this first sampling session followed immediately after the first session of instruction. For the other two learners, the first sampling sessions were 6 and 7 days after instruction. For the two instructed learners who chose to participate in two sessions of instruction, those tutorials were separated by 14 and 21 days. For the one instructed learner who chose to participate in three sessions of instruction, the first two tutorials were 6 days apart, with the third tutorial 16 days later. For all of these second and third instructional treatments, a sampling session occurred immediately after the instructional treatment. Therefore, for the three

instructed learners who chose to receive more than one session of instruction, their PostI phases included the sampling sessions after their second and third treatments.

4.1.3 The Data Collection Procedure

After collecting two preliminary samples of speech (a brief description of the learner and the learner's interests, and a spoken retelling of a legend each had invented and written for one of their classes) and a written sample (the legend), a standardized format was adopted, informed by the research discussed in the previous chapters. Approximately once every two weeks, each student provided three samples of language.¹⁴ All three samples involved the retelling of a short fable selected from a collection of Aesop's fables (*Fables of Aesop*, 1954), in spoken, written and cloze form. The selected fables are provided in Appendix C. The choice of the tasks is discussed in detail in Section 4.1.4.

Before each sampling session began, I greeted the learner and made the person feel comfortable with several minutes of informal conversation about topics such as school work and vacation plans. On three occasions during the second semester an additional task was completed before the fable retelling began. This task was a picture description task, developed for Chaudron and Parker (1990), and the data collected with this task was analyzed in Mellow and Fuller (1995). In addition, during the period of instruction, the fable retellings were usually done immediately after the instructional activity (on two occasions the instructional sessions were provided on a day between sampling sessions, as discussed in the previous section).

The procedures in the parallel sampling sessions are summarized below. First, at the beginning of each session, the student was given a sequence graphic (Mohan, 1986) that illustrated the events in a fable. Each graphic, titled only with the fable number, had five or six stages, with

¹⁴ During the first two sessions, the procedures were somewhat different. The first session involved an oral retelling followed by a cloze and a grammatical judgment and correction task, and the second session involved an oral retelling followed by a written summary. As discussed in Section 2.1, the judgment and correction task were analyzed in Mellow (1996) to determine if the students had difficulties with the ordering and non-iterability of functional categories in English (e.g., *John's the book, cf. Masterson, 1989).

each stage connected by an arrow, and with the sequence enclosed within a box. Each stage described events in the fable, using three to seven words, fairly basic vocabulary, simple non-past tense, very limited subordinate clauses, and almost no functional morphemes (e.g., only third person singular *-s*, plural *-s*; a few prepositions, pronouns, infinitival verbs, negatives, and conjunctions; but no articles, modals, complementizers, specifier words (e.g., *very*), etc.).¹⁵ Commas were the only form of punctuation that were used. An example of a graphic is provided in Appendix D.

Second, after clarifying any potentially difficult vocabulary, I read the fable aloud while the student followed the events on the graphic. During the reading, I occasionally answered questions that the learner asked about vocabulary, events, and causes. In addition, I elaborated on the text if the learner indicated confusion. Students occasionally made notes on the graphic, sometimes in Japanese, usually listing additional vocabulary items, and only on three occasions did students write a noun phrase with an article. Third, the student was then invited to ask any questions about the fable. Usually the student asked several questions to clarify the story or vocabulary. Fourth, when the student was fairly confident that the fable was understood, I read the fable aloud for a second time.

Fifth, the student took between 15 seconds and 2 minutes to prepare and to collect her or his thoughts. Then, often consulting the graphic, the student retold the fable and its moral into a microphone, usually at a language laboratory station, and always, except one occasion when a student was ill, in the language laboratory. While the student was retelling the fable, I sat beside

¹⁵ This graphic was provided to facilitate the learners' memories of the fable and provide consistent retellings across learners and sampling sessions. Articles were omitted so that the learners would not simply read and then use the target-like choice. Other grammatical morphemes were omitted to not draw attention to the absence of articles. However, the provision of this graphic may have modified the nature of these tasks, such that the learners were required to create a grammatical text from a non-target-like graphic, a transformation exercise that may have confounded knowledge of article use with oral and literacy skills. One apparent effect of this task, discussed in Mellow and Fuller (1995), was that more bare nouns (and fewer *a* overgeneralizations) were produced in the spoken and written retellings than were produced on a spoken picture description task that the learners completed on three occasion during the course of this study. Because the graphic was consistently provided during all retellings, comparisons of language use on these tasks are not affected by this factor. However, generalizability to authentic speech is not possible, an issue also discussed in sections 2.2 and 4.1.4.

the learner, nodding to indicate that I was listening and that I understood what the learner was saying. On a few occasions when learners struggled for quite a long period to find a word, I helped them, in order to keep the discourse relatively fluid and to reaffirm that there was an audience for this monologue.

Sixth, when the student had completed the recording (normally requiring just a few minutes), the student was given a lined sheet of paper and was asked to write a short paragraph summarizing the fable using the graphic. For this composition, the students were told that spelling was not important and that they did not need to use a dictionary. The students were not given a time limit for completion. The students usually took between 4 and 10 minutes and almost always used the graphic as a reminder. A sample of a completed composition sheet is provided in Appendix E.

Seventh, after the composition was completed, I gave the student a short cloze of the fable to complete. The cloze required six articles (three definite and three indefinite), as well as seven to nine other grammatical and vocabulary items which acted as distractors. The cloze task included blanks, both as words and suffixes, in which nothing was required for the passage to be target-like. These blanks were included so that the learners would not assume that a blank would require a lexical element. The students were not given a time limit for completion, but usually took between 2 and 5 minutes. The graphic was available for consultation but was not often used. A sample of the cloze task is provided in Appendix F. Eighth, after the completion of a session, the student was thanked and general arrangements were discussed, such as the date of the next sampling session or the time of additional consultation regarding course assignments.

4.1.4 Measurement Instruments

As discussed in sections 1.2 and 3.1, one important goal of L2 instruction is to enable learners to use the target language in spontaneous, authentic, non-classroom contexts. Because of the difficulties in obtaining authentic language (cf. the observer's paradox noted in section 2.2) and because of the importance of obtaining consistent and reliable measurements in order to assess

development (section 3.3.3.2), the measurement instruments were not chosen to elicit spontaneous speech, but instead were chosen for two related purposes. First, the spoken retelling task was chosen in an attempt to elicit language in which little or no attention was focused on the instructed elements, article use. Although not spontaneous, such a task may indicate the nature of the automatized knowledge that learners possess, knowledge that may be used when it is unlikely that they are monitoring their article use and when they have relatively limited attentional resources available. Second, a range of tasks was utilized to assess variation in interlanguage use, in particular, variations due to different degrees of attention to form.

As described above, a standardized, parallel sampling procedure was utilized. These consistent measurement procedures were used in order to increase reliability and hence to increase confidence in claims regarding development over time. Thus, all sampling sessions, except one, took place in the same language laboratory. I was the only interlocutor for all spoken retellings. Two pre-listenings and invited opportunities to ask questions about vocabulary and content were always provided. A similar sequence graphic was provided for each retelling.

In order to keep the linguistic demands of the tasks relatively parallel, eleven different Aesop's fables were chosen. As a narrative, a fable typically requires the use of articles to track the referents of non-generic noun phrases. The fables were chosen because the fable is a genre that the learners were familiar with. The learners said that they had been told fables in their childhood and had already written a legend for one of their classes, retelling it during one of the preliminary data collection sessions. In addition, fables usually have a fairly short, simple literal content that can be learned quickly, and they have moral implications that may be interesting, preventing the measurement sessions from becoming too boring or repetitive. In the final interviews, the non-native speakers reported that they enjoyed listening to the fables and discovering the conclusion and moral. Finally, because hundreds of fables exist, it was possible to select 11 relatively similar fables that are comparable in length, complexity, and content. However, the number of referents in each fable was not held constant, resulting in variations in topic predictability (cf. section 2.1.3).

In addition, variations in the presence of plural and mass nouns resulted in variations in the contexts for the indefinite article.

To control for the effects of possible differences in the specific fable versions, the order of presentation of the fable versions was partially counterbalanced. The order of presentation of the different fable versions is presented in Table 4.1 below.

Table 4.1
Order of Counterbalancing of Different Fable Versions

<u>Learner</u>	<u>Order of Fable Versions</u>										
T1	1	<u>2</u>	4	3	<u>5</u>	6	7	9	10	11	
T2	1	2	3	4	<u>5</u>	<u>6</u>	<u>7</u>	8	9	10	11
T3	1	2	3	4	6	5	<u>7^a</u>	<u>8</u>	9	10	11
T4	1	2	4	3	5	<u>6^b</u>	<u>7</u>	10	8	11	
C1	1	2	3	4	6	5	7				
C2	1	2	3	5	6	8	7	11			
C3	1	2	3	4	6	7					
C4	1	2	4	3	5	6	11				

Note. The sampling sessions that immediately followed instruction are underlined.

^a This learner received instruction one week before this session.

^b This learner received instruction immediately before and one week before this session.

The order of the first two fables was not counterbalanced in order to prevent the learners from realizing that they were not all performing the same tasks (subsequent counterbalancing was less noticeable because the scheduling of sampling sessions became more erratic in response to the learner's schedules). Initially, I was concerned that an awareness of systematic manipulations in the experimental design, including counterbalancing of fable version, would reduce the probability of unmonitored speech and would lead to increased speculation regarding the focus of the study. In addition, the order of fable versions was unfortunately not systematically counterbalanced during the final part of the study (cf. section 5.3). Although the order of presentation was not fully counterbalanced, the variation in presentation order should have partially offset systematic changes due to differences in versions of measurement instruments. In particular, the first fable version after instruction varied considerably, with either version 5, 6, or 7 being used during those

sampling sessions. In addition, due to variations in scheduling, the calendrical date at which each fable version was used also varied considerably.

As noted above, in each session three different types of retellings of the same fable were used in a fixed order: spoken, written and cloze. These types of tasks were chosen in order to hold certain factors constant in all tasks and yet vary other factors systematically between the tasks. Many factors have been proposed to account for aspects of systematic language variability (e.g., Wells, 1986, p. 110; Young, 1989, p. 75) (cf. section 2.2). Two of the factors which were held constant in the three different types of retellings are discussed in detail: genre and communicative pressure.

Each of the three tasks was held constant in terms of genre: each is a narrative (Hatch, 1992, pp. 165-174). Narratives typically utilize certain syntactic and discourse features and hence provide a specific type of linguistic context (cf. Young, 1989, pp. 83-4). Of particular interest for this study is that the tracking of specific referent noun phrases is an important aspect of narratives. In Tarone (1985) and Tarone and Parrish (1988), the genre was not controlled for in the different tasks that were used to study variability of article use and this may have resulted in some of the variation in the results. Thus, the surprising result of Tarone and Parrish (1988, pp. 34-5), that accuracy was worse on grammaticality judgments than in spoken narratives and interviews, may have been due to the absence of extended, cohesive discourse in the grammar test. The desire for extended, cohesive discourse motivated the choice of the cloze test as the task in the present study that has the greatest focus on form.

Each of the three tasks was also held constant in terms of communicative pressure (cf. Tarone & Parrish, 1988, pp. 34-35). Communicative pressure reflects the necessity of the correct use of language in order to avoid substantial confusion and misunderstanding on the part of the listener. Although communicative pressure may be a desirable trait for purposes of authenticity, the three tasks used in this study all involved a relative absence of communicative pressure: I, as listener, already knew the fable. Communicative pressure was omitted for all retellings in order to keep this property constant for all tasks. The alternative, to have communicative pressure for all

tasks, was determined to be impractical and undesirable because it would require that either a new listener or reader would be required for each task (someone who presumably did not know the fable), or that a completely new fable would be required for each of the three types of fable tasks (and two different native speakers would have told and listened to each of the fables). This change of interlocutor or fable version might have seriously reduced the comparability of the tasks, for example, in regard to the status and gender of the interlocutor, or the topic and complexity of the fable. Such a procedure would also have made each sampling session very long and arduous, likely causing even greater attrition in the volunteer participants. The tasks in this study did involve some amount of communicative pressure, that is, the pressure that comes from the assumed ritualized cultural behavior of role play, acting out or repeating a known situation or story. This type of activity is quite common in school and in the final interview the learners reported that they were comfortable and even felt natural while reciting these fables to a listener/reader who knew the plot and moral.¹⁶

Although certain factors were held constant in the tasks, the purpose of this study was to assess variability in interlanguage use and therefore required the elicitation of language use under specific varying conditions. The different types of retellings were chosen so that they would vary according to four properties or parameters. These four parameters of variation are summarized in Table 4.2.

The first parameter that varied in the tasks was the degree to which a learner was required to focus on or attend to specified aspects or forms of language. The cloze test focused the learner's

¹⁶ Researchers have proposed a number of principles and guidelines for the construction of tasks for which speakers will have communicative pressure and therefore will be unlikely to pay attention to speech (e.g., Brown & Yule, 1983b, pp. 102-149; Labov, 1972, pp. 207-216). Such principles were not utilized in the present study because it was believed that it would be difficult to devise 11 parallel versions of these tasks for the different elicitation sessions and because it was believed that it would be difficult to devise three parallel tasks within each elicitation session if one task involved communicative pressure and the other two tasks did not involve communicative pressure.

Table 4.2
Four Parameters of the Three Retelling Tasks

<u>Retelling Task</u>	<u>Parameters of Variation</u>			
	<u>Attentional focus</u>	<u>Time</u>	<u>Modality</u>	<u>Prior Attention</u>
Spoken	- attentional focus	- time	speech	limited (1st recall)
Written	- attentional focus	+ time	writing	moderate (2nd recall)
Cloze	+ attentional focus	+ time	writing	considerable (3rd recall)

Note. Plus = [+]; minus = [-].

attention on the item that belonged in a blank, with 6 articles required in the 13 - 15 blanks of each cloze test in this study. In contrast, the spoken and written retellings did not focus the learner's attention on any specific forms. Instead, these tasks required the learner to consider a large number of factors, including aspects of pronunciation or spelling, grammar (e.g., tense, article suppliance, prepositions), vocabulary choice, discourse organization (e.g., rhetorical structure, cohesion, NP tracking), and sequential content. To reduce the burden of recall and encourage reasonably fluent narratives, the sequential graphic that was given to the learners provided some content-oriented support. This absence of a focus on specific forms was maintained by a strict ordering of the tasks, with the spoken retelling always first, followed by the composition and then the cloze task last. If the order had been reversed (or counterbalanced), during the spoken and written retellings the learner might have recalled the items which were attended to while completing the cloze. In addition, because a focus on specific forms is possible in the composition, if the order had been reversed, the forms attended to while composing might have influenced the learner's attention during the spoken retelling.

The amount of time that a learner had to attend to or focus upon each aspect of the task was the second way in which the tasks varied (cf. planning or processing time: Ellis, 1989, pp. 33-4, 40-1). The oral fable retelling did not allow the learner very much time to attend to the numerous specific aspects of the task. These spoken monologues encouraged the speaker to maintain a moderate pace so that the discourse was fluent and understandable. Although there are situational and cultural influences on acceptable pause length, the conventions of fluent storytelling pressured

the learners to limit the length of pauses. These time restrictions were very important because the attentional resources of speakers are quite limited. As a result, the spoken task reduced the planning and monitoring of specific aspects of the output: only a very small number of items could be attended to. For a task that had never been performed before, the retelling of a newly learned fable, attentional resources may have been used to focus upon the content and very few resources would have been left for attending to linguistic form. In contrast, the written and cloze tasks had no time restrictions. As a result, the written and cloze tasks provided the opportunity to attend to form and use conscious rules in choosing morphemes, words, sentences, and discourse structures. As noted above, only the cloze task required attention on specified forms. Thus, the composition allowed a focus on forms, but the learner decided which of the many aspects of language were consciously considered.

The *time* factor was confounded with a third factor, modality: speech vs. writing (cf. Crookes, 1988, p. 6 for a related discussion). In fact, the pressure of maintaining fluent speech probably was the main reason why the oral retellings had restricted time available for attention to specific aspects of language. In addition, the process of writing (e.g., planning) may promote attention to various aspects of linguistic form, and written words (provided or composed) may serve as reminders of aspects of language that a person may focus upon. If the process or artifacts of writing cause attention to form, then these aspects of the writing would support the claim that the composition and cloze tasks encouraged more attention to form. However, modality involves other properties that may have confounded the distinctions drawn in Table 4.1. For example, if the learners were not equally literate, then their writing ability may have affected their performance on various specific aspects of the tasks. In other words, the modality of the tasks may have confounded the results. However, at least for this one aspect of modality (i.e., literacy in the mother tongue), there is some evidence that writing ability does not affect performance on linguistic measures. Cumming and Mellow (1996) found that, in the compositions of 24 adult Francophone students, writing expertise in the mother tongue did not correlate with accuracy of L2 article use. Thus, although certain aspects of modality either support or may not affect the claimed

task distinctions, because these two factors are confounded, the results must be examined carefully and interpreted cautiously.

The fourth factor that differed between the tasks was a factor that might be referred to as the amount of prior attention given to the execution of the task. Planning (e.g., Crookes, 1988, 1989) would be one aspect of prior attention to a task. A second type of prior attention is what might be called rehearsal: the number of times that the task has been practiced or listened to prior to the specific performance. Rehearsal may not be an exact duplication of all conditions, and therefore listening to the fable twice may have constituted prior attention for the spoken retelling, the spoken retelling may have constituted additional prior attention for the composition, and the composition may have constituted additional prior attention for the cloze. In the rehearsals of the fable, attention may not have been deliberately focused on specific aspects of the task, such as article use, but difficulties in the execution or comprehension of the task may have resulted in attention being drawn to specific aspects of language. As a result of the fixed order of the fable tasks, spoken retellings had the least prior attention, clozes had the most prior attention, and compositions had an intermediate amount.

These four task parameters allow a task classification according to the probability of learners attending to form and using conscious grammar rules during a given task. The tasks have been ranked in Table 4.3 according the total number of properties that permit or encourage attention to form and use of rules.

Table 4.3

Total Properties of Tasks that Contribute to Attention to Form and Use of Rules

<u># of Properties</u>	<u>Retelling Task</u>	<u>Specific Attention</u>	<u>Time</u>	<u>Writing</u>	<u>Prior attention</u>
1	spoken	no	no	no	a little
3	written	no	yes	yes	moderate
4	cloze	yes	yes	yes	considerable

If these psycholinguistic processing distinctions are valid, then they suggest that spoken retellings were least likely to encourage attention to form and use of rules, cloze tasks were most likely to

encourage attention to form and use of rules, and written retellings may have encouraged an intermediate amount of attention to form and use of rules.

These task differences allowed an examination of variability of language use, as predicted in Hypothesis 1 in section 2.5. In addition, the use of these different tasks permitted an examination of the extent to which language knowledge became automatized, that is, was accessed when attentional resources were likely focused elsewhere during the spoken task. Finally, the factor of available time was also important because of the cognitive and linguistic issues raised in section 2.4. It was hypothesized that the efficient deletion of the redundant presence of *the* would occur in tasks with relatively limited attentional resources. Therefore greater omission of *the* was expected on the spoken retelling. In contrast, it was hypothesized that priming due to the redundancy of *the* would occur on tasks with relatively abundant attentional resources. Therefore, the facilitation of priming was expected on the written retelling and on the cloze.

In sum, these language tasks did not elicit the natural spontaneous language use that may be of greatest interest to educators, but nonetheless may provide considerable insight into language use and development. In particular, the tasks (and their order of completion during each session) were designed to control or eliminate certain aspects of variability that were not investigated in this study. In addition, the tasks were designed in order to investigate other aspects of variability, those associated with psycholinguistic processes such as attention to form, priming, and the efficient omission of redundant elements.

4.1.5 Instruction

After approximately 90 days of the study, I offered sessions of instruction regarding English articles. As a result of the learners' diverse and busy time schedules, I administered the intervention individually for each learner, with each session generally lasting between 45 minutes and one hour. As noted above, the eight learners were offered as many as three sessions of instruction: one learner chose to have one session, two learners chose to have two sessions, and one learner chose to have three sessions.

The purpose of this study was not to compare different types of instruction. Instead, the purpose was to assess the potential effects of relatively optimal instructional sessions. In order to design instructional activities that would have been most likely to result in development, the instruction that was provided to the learners was based upon several specific suggestions that emerge from different theoretical accounts of the nature of language learning. The instructional activities included a presentation of a grammatical explanation, followed by three types of practice activities: two kinds of input processing activities (referentially-oriented and affective), followed by an output practice (strategy use) activity. After each aspect of the instruction is described in this section, the section concludes with an explanation of the contextualization of instruction that was employed to avoid the Hawthorne Effect.

Grammatical rules for English article use were presented only to the instructed learners at the beginning of the first session of instruction in an explicit, metalinguistic, and deductive manner. The presentation was intended to raise the learners' consciousness about this form (cf. Sharwood Smith, 1981) and to ensure that they had been supplied with this grammatical knowledge. The specific form of the rules and their presentation was based upon Celce-Murcia and Larsen-Freeman (1983) and Master (1990, 1994), and was a revised version of a grammatical presentation that I had used one year earlier with a similar group of L2 learners as part of a preliminary pilot study of instructed ESL article development. The initial presentation of the grammatical rules involved a detailed discussion of a two-page handout (Appendix G). At the beginning of the second and third sessions of article instruction, the handout was briefly reviewed, emphasizing article use with referentially definite NPs.

The practice activities focused on a subset of English article use. Because it was believed that the entire article system was too complex to be presented in sufficient detail during a relatively short period of instruction (cf. the issues discussed in Master, 1994), and because of the desire to study only certain aspects of English article use, only a subset of the English article system was chosen for the practice activities. This subset included *the*, *a* and \emptyset used with referentially definite NPs, primarily using singular count nouns.

The learners were provided with two types of input processing activities, based upon VanPatten (1993, 1994). Input processing activities are designed to cause changes to learners' interlanguage knowledge by drawing their attention to specific features of grammar in the input, thus ensuring that the language elements and their meanings are comprehended and that correct form-function mappings are processed as part of the intake to the developing system. VanPatten (1993) distinguished between two types of input processing activities, referentially-oriented activities and affective or learner-centered activities, suggesting that referential activities should precede affective exercises.

In referentially-oriented activities, learners use an immediate concrete reference, such as a picture, to ascertain the truth value of a sentence. The learner must respond to the input in a manner that indicates comprehension of the language element. VanPatten (1993) suggested that such practice may begin with a written activity in order to facilitate input processing of particular linguistic features.

Three sets of referentially-oriented activities were developed for this study, with a different set used during each of the three instructional sessions that were offered to the learners. There were eight activities in each set, with one activity per page. In order to insure the comprehension of the form-function mappings involved with the tracking of NPs, each activity included two pictures in a sequence on a page, and the learner was required to determine which of two sentences correctly described the second picture. In order to choose the correct sentence for the second picture, the learner needed to (a) determine whether the referent in the second picture was new, and (b) attend to and comprehend the meaning of the article in each of the alternative sentences.

For each set of referentially-oriented input processing activities, the students were given the following instructions:

For each of the following pairs of pictures,

- 1) Look at the first picture and read the sentence that is below it; and then
- 2) Look at the second picture and choose the sentence that best describes the picture.

An example of one of the eight activities in a set was two cartoon pictures of the same dog (i.e., Snoopy, from Schulz, 1973). In the first picture (on the top of the page), Snoopy is eating from a bowl and the following sentence is printed below the picture: *A dog is eating from a bowl*. In the second picture (on the bottom of the page), Snoopy is cooking pancakes and there are two sentences printed below the picture: *The dog is cooking pancakes.* and *A dog is cooking pancakes.* In this example, the correct sentence for the second picture is *The dog is cooking pancakes* because the referent (Snoopy) is the same in both pictures. The student therefore checked the box in front of the sentence that was chosen or respond orally to me. In each of the three sets of items, four of the second pictures had a new referent and four of the second pictures had an old referent, presented in a random order. The sentences were monoclausal, with action verbs. In the first set, the referent in question was always in the subject position of the sentences. In the second and third sets, the referent was equally distributed in the subject and non-subject positions of the sentences.

The second type of input processing activities are affective activities. These activities also require the learner to respond to a question or stimulus in a manner that indicates that the language element has been comprehended. However, rather than using an immediate concrete reference, these activities require the learners to draw upon their own personal knowledge, asking for personal opinions, responses, and inferences (VanPatten, 1993, pp. 439, 441).

Three affective activities were developed for this study, with a different activity used during each of the three instructional sessions that were offered to the learners. In the first two activities, there were 25 sentences in a narrative. In the third activity there were 10 sentences in a narrative. In order to insure the comprehension of the form-function mappings involved with this aspect of article use, the learner was required to determine the truth of each sentence, based upon the referents that had appeared previously in the discourse and based upon the learner's inferences and knowledge of the world. In order to determine whether each sentence was correct, the learner had to (a) attend to and comprehend the meaning of the article(s) in the sentence, and (b) determine whether the referent of the noun phrase was new. A subset of the sentences from the second affective activity is provided in Figure 4.2. In this example, the anticipated response for sentence 2

Going Downtown - Guess which of the following things happened to Dean when he went downtown last Saturday. Put 'yes' or 'no' in the blank. We will discuss them when you are done.

- ___ 1. A #7 bus picked him up one block from his house.
 ___ 2. A #7 bus took him downtown.

Downtown, many people were playing instruments and singing.

- ___ 3. A man played a violin.
 ___ 4. A man sang "Heartbreak Hotel" and played a guitar.
 ___ 5. The man played a flute.
 ___ 6. The man played "Amazing Grace" on a flute.
 ___ 7. Some people gave money to a man who played the flute.

Figure 4.1 Sample activity of affective input processing instruction, excerpted from the second exercise

is *no* because the presence of A #7 bus implies that after one #7 bus picked Dean up, it was a new bus that took him downtown. Instead, one would expect that Dean rode only one #7 bus and therefore the expected sentence would be *The #7 bus took him downtown*. The anticipated answer for sentence 5 would probably be *no*, because the presence of The man implies that it is the man in sentence 4 that is playing a flute (and singing "Heartbreak Hotel"). However, a *yes* interpretation is possible if the learners inferred that the same man had stopped singing and had started playing a flute. In this way, the activity required the learners to focus on the meaning of the articles in the written input, drawing upon their knowledge of the world, and inferring whether each referent was new. In each of the three activities a mixture of sentences was provided, with between 68% and 80% of the sentences possibly having a *yes* interpretation. The sentences were generally short (no more than 16 words in length) and the referents that could result in a *no* response occurred in either subject or non-subject position.

After the two types of input processing activities, the final instructional activity in each session was an oral production activity in which the learners were asked to describe a series of

events that had happened to them or in which they had been involved. Prior to recalling these events, the learners were asked to pay attention to their use of articles when they described their experiences. These activities involved a metacognitive strategy of selective attention during self-monitoring of spoken output (cf. O'Malley & Chamot, 1990, esp. pp. 44, 180, 182; O'Malley et al, 1985). An oral production activity involving strategy use was chosen because of the apparent effectiveness of strategy training (e.g., O'Malley et al, 1985) and because of the possibility of learning through comprehensible output (Swain, 1985, 1993). Swain (1993) discussed a number of ways in which producing language may result in language acquisition, suggesting that (a) the meaningful practice of one's linguistic resources may result in the development of automaticity of use, and (b) production may force the learner to move from semantic processing to syntactic processing, permitting attention to redundancies that may be ignored during comprehension (cf. Gary & Gary, 1981, p. 3, quoted in Swain, 1993; VanPatten, 1993; Mellow & Cumming, 1994).

In the oral production activity that was provided in the instructional sessions, the learners were allowed to choose the events that they described. The learners chose to describe either (a) trips that they had taken during Christmas vacation, to various locations in Canada and the United States, or (b) examples of days or sessions of their recreational activities, including skiing and skating. These oral narratives varied in length, five lasted for 10 to 15 minutes, one lasted for 30 minutes, one lasted for an hour, and one last for an hour and a half. Although I (as both the researcher and the instructor) generally behaved as a conversational partner during these activities, responding to the content of narrative and often requesting clarification and elaboration regarding the learners' experiences, on occasion (no more than once or twice in five minutes) I noted and corrected non-target-like use of articles.

As noted in sections 3.3.2 and 4.1.1, a number of strategies were employed to avoid the Hawthorne Effect, that is, to prevent the learners from knowing that article use and development was being studied. Two of these strategies, (d) and (e) in section 3.3.2, were used to avoid special emphasis being given to instructional treatment: the provision of many other types of non-

compulsory instructional activities during the period of the study; and the justification of the provision of the experimental treatment as a response to the requests of the learners.

With respect to the first strategy related to instruction, in exchange for the many hours during which the learners provided data samples for this research project, I provided them with various types of instruction and academic assistance with their course work. At the beginning of the study, the learners were offered a series of five one-hour seminars that dealt with topics in which they expressed an interest: (a) learning to use the Macintosh computer; (b) register: using appropriate formal and informal language (two seminars); and (c) rhetoric: organizing essays and papers according to North American university standards (two seminars). When the students became too busy to continue these seminars, for the remainder of the school year they were offered assistance during office hours in the computer room / language laboratory in the basement of their residence building. Therefore, throughout the duration of the study, I provided individual tutoring regarding many aspects of academic writing. This helped the learners in three ways: first, by keeping the computer lab open for additional hours so that they (and other ESL students) could write their course papers; second, by providing times in which I could help them learn to revise and edit their papers; and third, by providing them with an opportunity to speak English with a native speaker who expressed an interest in their school work, holidays, future plans, etc. During this personal tutoring, I did not provide any special attention to article use, attempting to avoid additional instruction in article use. The two groups participated fairly equally in the office hours sessions, with two instructed learners dropping by most regularly, three control learners dropping by occasionally, and the remaining two instructed learners and one control learner rarely seeking additional tutoring. During these same office hours, students would come in and provide the language samples which form the basis of this study.

The second strategy for avoiding the Hawthorne Effect was to explain to the learners that the instruction regarding article use was part of this continuing academic assistance and had been requested by the learners. This was not a misleading explanation because, during tutoring regarding the revision of their papers, several students had told me of their difficulty with articles

and asked for assistance with articles. Although I was selective about which requests for assistance led to the provision of instructional materials, the treatment was nonetheless requested by the learners. This instruction regarding articles was offered to the students in a manner similar to the offering of the seminars. I intended to provide similar practice for verb tense choice, another problem area raised by the learners, and announced this intention to the learners, but practical considerations did not permit time for this activity.

4.2 Data Analysis

4.2.1 Transcription and Coding

I transcribed the spoken data from audiocassettes, paying particular attention to the elements in pre-nominal position of referentially definite NPs, NPs that refer to specific referents and whose referents are assumed to be known to the hearer. Thus, generics and nonreferentials were not coded. Because each fable had a one or two sentence moral that was comprised almost entirely of generic NPs, the moral of each fable was excluded from the analyses. In addition, referential indefinites (e.g., the first mention of a specific referent) were coded but not included in the analyses. Furthermore, this study was only concerned with article use (*the*, *a* and \emptyset): other possible pre-nominal elements were not considered in this analysis, including those elements which may have the same function as articles, for example, *this*, *that* and *one*. Mellow and Fuller (1995) provided an analysis of the use of these elements by these learners on a different task.

Particular attention was paid to the transcription of four categories of pre-nominal items, each of which was transcribed with distinguishing symbols:

- (a) those judged to be articles, which were transcribed as *the*, *a* or nothing (the majority of the items; \emptyset was not marked on the transcripts);
- (b) those judged to be hesitation syllables, which were transcribed in a rough phonetic transcription as *uuh*, *uum*, *mmm*, *hmm*, *uur*, *uun*, *ooh*, (there was a large number of these, especially for certain learners; phonetically, the first item is somewhat similar to *a*);
- (c) those found to be indecipherable (there was a relatively small number of these); and

(d) those that were somewhat ambiguous, but for which one interpretation appeared most likely, which were transcribed as indicated below and placed in parentheses (there was a small number of these), for example,

(i) (*the*) and (*a*); probably *the* and *a*;

(ii) (*the a girl*); probably a change from *the* to *a* in mid-speech, but potentially *the* followed by a hesitation syllable;

(iii) (*and a boy*); probably *and* followed by *a*, but potentially (a) [ænd da] *and the*, with the alveolar stops merging in fast speech; (b) [ænd ʌh] a hesitation syllable; or (c) [ændə] an epenthetic reduced vowel (schwa) used to preserve CVCV phonotactic patterning;

(iv) (*aaa*), phonetically [æææ]; probably a hesitation syllable, but potentially a lengthened indefinite article; and

(v) (*aah*), phonetically [ʌʌʌ]; probably a hesitation syllable, but potentially a lengthened indefinite article;

A second transcriber, a native speaker of English who was an experienced ESL teacher and had worked extensively with Japanese learners, then carefully listened to the tapes while reading the transcriptions. The second transcriber noted any differences of opinion for the first three categories and either confirmed or rejected the transcription for the fourth category. Noun phrases were eliminated from the analysis if there was doubt that a pre-nominal word was an article. Consequently, the following items were retained for the analyses: those in category (a), and those in category (d), such as (i), (ii), and (iii), that were confirmed as specific articles by the second transcriber. The compositions, which had been handwritten, were typed into a word processing package so that they could be coded in a similar fashion to the transcript.

The spoken and written data were then coded by two raters: myself and a native English speaker who was an experienced ESL teacher. Before coding the corpus, the two raters trained together on and discussed about 5% of the fables. The data were arranged so that there was one clause per line, with a clause defined as any verb, finite or non-finite, and its subject, complements

and modifiers, even if the subject or complements were missing. The only exception to this definition was a series of two verbs without any intervening NPs, for example, *fox wanted to get his meat* or *a fox tried to hmm trick him*, with the series of verbs underlined (cf. subjectless non-finite clauses as direct object: Quirk & Greenbaum, 1973, p. 361). These constructions with a series of verbs were categorized as a single clause. Examples of these clausal classifications are provided in 20), 21) and 22) below. Using colored pencils on print-outs of the word-processed data, each prenominal context for each referentially definite noun was coded for three characteristics: (a) presence of *the*, *a* or zero; (b) syntactic position: subject or non-subject position; and (c) distance since previous mention of that referent.

The following conventions were adopted for coding the first category, the presence of *the*, *a* or zero. First, meta-discoursal phrases were not coded, for example, *the story is about* and *once upon a time*. Second, for the spoken data, in a series of prenominal elements with more than one article, only the last article was coded, and articles were coded even if there was a hesitation syllable intervening between the article and head noun. For example, both *a the man* and *the uuh man* were coded as *the*. Third, in co-ordinated NPs, both prenominal contexts were coded, even though the second context can be bare for native English speakers, for example, *The horse and Ø donkey carry the luggage*. The combination of *the Noun* followed by *Ø Noun* was quite rare in the data. Fourth, for the spoken data, in a series of repeated elements, NPs, NP + Verbs, or clauses, only the last phrase in the series was coded. For example, only the underlined NPs in the following utterances were coded (+ indicates a short pause):

- 20) so mouse + mouse capture + mouse capture the lion
- 21) so mouse come to + mouse came to lions + uuh mouse came to + came to something near near the lion (= 1 clause)

If, in repeated clauses or clause fragments, there was a change of verb, then the clauses were considered to be distinct and coded as such, as in 22) below, with only the underlined NPs being coded:

- 22) uuh and fox saw +
fox wanted to + wanted to get the meat

The following conventions were adopted for coding the second category, syntactic position: subject or non-subject position. Any NP preceding the verb was considered to be in subject position. Any NPs following the verb were considered to be in non-subject position. A further distinction was made for the non-subject positions, between direct object and object of a preposition or indirect object. These distinctions led to difficulties distinguishing phrasal verbs from prepositions in the interlanguage data. However, this distinction was not used in this analysis. In addition, there was a very small number of prepositional phrases at the beginning of clauses. These were excluded from the present analysis.

The following conventions were adopted for coding the third category, distance since previous mention of that referent. Current referents were those referents that had been mentioned in the previous clause. Recent referents were those referents which were last mentioned two to five clauses earlier in the discourse. Old referents were those referents which were last mentioned more than five clauses earlier in the discourse. New referents were those that had not been mentioned previously in the data. On a very small number of occasions, the learners switched from narrating the events in the third person to providing a direct quotation of speech in the first person. These quotations were considered to be part of the continuing discourse, even if the referent was referred to as *I* or *you*.

These conventions are a crude measure of referential distance, for at least two reasons. First, referential distance was quantified in terms of clauses rather than elapsed time. In particular the presence in the spoken data of intonation units and pauses of differing lengths could significantly affect the speaker's perception of the distance or time since the last mention of a referent. However, using the clause as unit did allow the same unit to be used in both the spoken and written data. A second concern was that it was the speaker's perception of the need to emphasize the re-introduction of a referent that influenced the choice of forms. Therefore, an ideal

measure would have assessed how speakers used forms in response to their perceptions, rather than using a simple quantification of distance in terms of number of clauses.

In addition, the category of new referents was particularly complex. Indefinite first-mentioned NPs can be introduced with either *a*, for singular count nouns, or with \emptyset , for plural and mass nouns. However, it is difficult to know whether the \emptyset article in *he got uuh \emptyset salt in his back* is due to target-like knowledge of article use and the mass status of the noun, or due to omission of an article. Equally, it is difficult to know if the absence of an indefinite article with a count noun is due to the non-target-like assumption that the noun is mass. Finally, often it could have been assumed that the learner had adopted a story-telling perspective and assumed that the fable should be told as if the listener knew nothing about the fable. However, because the learner could have assumed that I, as listener or reader, knew the fable, having twice told it to the learner, the use of a definite article with a new referent could be considered target-like. For this reason and others discussed in section 2.1.3 (e.g., the relative difficulty of distinguishing *aaa* and *uuh*), the category of indefinite new referents was not considered in this study.

However, definite first mentioned NPs were included in the analyses. A number of first-mentioned referents could be assumed known to the listener or reader because of entailment from the previous discourse (*he eats things and give the leftover to charity; he put the fox's tail to a fire and then the fox ran away because of the pain*), post-modification (*mouse runs over the body of a sleeping lion*), uniqueness in the world (*the mouse is running on the ground*), or other reasons. There were 155 definite first-mentioned NPs in the corpus (8.4% of the 1844 definite contexts). Definite first-mentioned NPs were included in the analyses for the first of the five dependent variables described in section 4.2.2 below.

Finally, there were very few referents that were coded as old according to the criteria of *five clauses since last mention of the referent* (116 instances or 6.3% of the corpus of 1844 definite NPs, averaging less than one, 0.86, per fable). As a result, this category of topic continuity was unfortunately not included in this study.

When slight differences in directions for coding were taken into account, the two raters achieved 96.8% agreement on 5550 judgments. The differences in coding were mainly caused by differences in judgement of the distance to the previous mention of the referent (a mention was sometimes missed when it was a pronoun, a synonym, or a possessive rather than a full NP), by miscoding a series of repeated prenominals or NPs, and by noun phrases that were missed by the first coder and which were each considered to be three differences in coding, one for each of the three characteristics. After verification, the codings were transferred to scoring sheets, with one sheet per fable (a sample scoring sheet is provided in Appendix H).

4.2.2 Analysis of the Data: Variables and Summary Measures

4.2.2.1 Definitions of the Five Dependent Variables

The data from the scoring sheets were used to develop scores for each learner on five dependent variables. As discussed in detail in section 4.2.2.2 below, the scores on these five dependent variables were analyzed in four different ways in order to assess the development of the learners.

Providing an overall indication of the provision of the definite article for referentially definite (RD) NPs, the first dependent variable considered in this study was *the* suppliance for referential definites (abbreviated Def). This value was computed according to the procedure in 23) below.

$$23) \frac{\text{the number of RD NPs for which } \textit{the} \text{ is provided}}{\text{the number of RD NPs for which } \textit{the}, \textit{a} \text{ or } \emptyset \text{ is provided}} \times 100$$

For RD NPs, *the* is the only article which is target-like. In the corpus, there were only 21 instances in which *a* was overgeneralized to referential definites (1.1% of the 1844 item corpus): Most of the non-target-like forms were bare nouns. For this variable, RD NPs include those NPs that were definite on first mention, for reasons such as entailment from the previous discourse, post-modification, uniqueness in the world, as discussed in section 4.2.1. This first dependent variable, providing an indication of *the* suppliance, but not of the overgeneralization of *the*, is

similar in scope to supplied in obligatory context (SOC) analyses provided by other researchers (cf. Pica, 1983a; Master, 1987; Cumming & Mellow, 1996; cf. section 2.1.2).

In order to analyze redundancy and polyfunctionality mappings, the next four dependent variables referred to specific linguistic contexts in which RD NPs appear. For these four dependent variables, the overgeneralization of *a* (i.e., the substitution of a non-target-like form) and NPs that were definite on first mention were not considered because the research questions were primarily investigating the conditions under which *the* was omitted, for reasons of efficiency or the absence of additional facilitation, including the factor of referential distance. Because the number of *a* overgeneralizations and the number of NPs that were definite on first mention was relatively small (totally 9.5% of the corpus), this aspect of the dependent variables eliminated relatively little information about the use of articles by these learners (for a discussion of the overgeneralization of indefinite markers by these learners on a different task, see Mellow & Fuller, 1995).

The second dependent variable, *the* suppliance in subject position (abbreviated Sub), was computed according to the procedure in 24) below.

$$24) \frac{\text{the number of RD NPs in subject position for which } the \text{ is provided}}{\text{the number of RD NPs in subject position for which } the \text{ or } \emptyset \text{ is provided}} \times 100$$

The third dependent variable, *the* suppliance in non-subject positions (abbreviated NSub), was computed according to the procedure in 25) below.

$$25) \frac{\text{the number of RD NPs in non-subject positions for which } the \text{ is provided}}{\text{the number of RD NPs in non-subject positions for which } the \text{ or } \emptyset \text{ is provided}} \times 100$$

The fourth dependent variable, *the* suppliance for current NPs (abbreviated Cur), was computed according to the procedure in 26) below.

$$26) \frac{\text{the number of current RD NPs for which } the \text{ is provided}}{\text{the number of current RD NPs for which } the \text{ or } \emptyset \text{ is provided}} \times 100$$

The fifth dependent variable, *the* suppliance for recent NPs (abbreviated Rec), was computed according to the procedure in 27) below.

$$27) \frac{\text{the number of recent RD NPs for which } \textit{the} \text{ is provided}}{\text{the number of recent RD NPs for which } \textit{the} \text{ or } \emptyset \text{ is provided}} \times 100$$

As in Ellis (1990, p. 317) and Mellow and Cumming (1994, pp. 460-461), a score for each sampling session (i.e., either a written and spoken fable retelling) was reported only if the learner provided a minimum of three contexts for a dependent variable, for example, three subject positions in which *the* was required. In the cloze tasks, only three contexts for *the* were present in each task¹⁷. Calculations based on such a small number of contexts reduced reliability because the suppliance of one more or one less article could have caused a shift of as much as 34% in the session score. However, if the minimum number of occasions had been set higher, additional data would have been lost. Even with the minimum number of contexts set at three, 90 session scores (13.3% of the scores) were not reported. Of the remaining 585 scores that were used in the analyses, 48 (8.2%) were based on three contexts, 69 (11.8%) were based on four contexts, 60 (10.3%) were based on five contexts, 216 (36.9%) were based on six to nine contexts, and 192 (32.8%) were based on 10 or more contexts.

4.2.2.2 Descriptions of the Four Summary Analyses

The scores of the learners on these five dependent variables were summarized using four different methods: (a) group mean phase scores; (b) inferential statistical tests; (c) phase regression lines; and (d) line graphs. Each of these summary analyses was used to make different types of comparisons required by the 12 research hypotheses listed in section 2.5. The first three summary analyses are used to make comparisons between groups and within groups, and across time (preintervention and postintervention phases), tasks, and linguistic contexts. The fourth summary analysis, line graphs, is used to make comparisons within individuals across time, that is, to assess the developmental course of SLA (i.e., Hypothesis 12).

¹⁷ Unfortunately, one cloze task only had two required contexts for *the* and therefore was not considered in the data.

The choice of these different methods of summarizing the data reflected an attempt to optimize the amount and type of information extracted from the scores, while also maintaining reliability in spite of the considerable (although perhaps not unusual) variability in these interlanguage data. For the scores on a dependent variable in either the preintervention (PreI) phase or the postintervention (PostI) phase, two types of information were sought: central tendency and trend. Central tendency in a phase was desired in this experimental study so that a comparison could be made between typical scores from phase to phase, or dependent variable to dependent variable. To indicate a central point in the distribution of the scores for each phase, group mean phase scores were used. Parametric inferential statistical tests, based upon the group mean phase scores, provided a more powerful comparison of the central tendencies of the groups and of the different dependent variables. In addition, non-parametric inferential statistical tests, not based upon group mean phase scores, were also used in this study to compare group scores. The second type of information sought in this study was trend, the possibility of a systematic change in the scores over time. Two methods were used to provide information regarding the trend(s) in the data: regression lines for each group for each phase, and line graphs for each individual for each phase. The four summary analyses for central tendency and trend are each described in detail.

4.2.2.2.1 Group Mean Phase Scores

The first two types of analyses, group mean phase scores and inferential statistical tests, assessed central tendency in a phase and were computed using a *pooled phase score for each individual* for each dependent variable. Two methods of determining a phase score for each individual for each dependent variable were considered: (a) computing a suppliance score for each sampling session and then averaging all of the session scores to create a mean phase score for each individual, and (b) pooling all of the instances of and contexts for the suppliance in a phase in order to create a pooled phase score for each individual (cf. Mellow & Cumming, 1994, p. 461). The first option, a mean phase score for each individual, was not used in this study because the small numbers of contexts for a dependent variable in some sampling sessions may have either

eliminated that information or may have distorted the phase score because sessions with small numbers of contexts provide a fairly crude score (cf. 4.2.2.1 and the computation of the regression lines, discussed below).

As a result, the second option was used for the measure of central tendency: all of an individual's scores in a phase (PreI or Post I) were pooled for each dependent variable using the formulas in 4) through 8) above. Cook and Campbell (1979, p. 43) recommend the use of more aggregated units to increase the reliability of a measure and this pooling of an individual's scores for a phase presumably resulted in a more reliable phase score for each individual. This concern for reliability appeared to be particularly important because of the variability in the scores (high standard deviations and low r^2 values were typically reported in the results in Chapter 5), in spite of the attempt, discussed in section 4.1.4, to use parallel versions of the testing instruments (cf. section 3.3.3.2). The pooled phase scores for each individual for each dependent variable are reported in Appendix I. The pooled phase scores for each individual were used to compute the non-parametric inferential statistical tests discussed below. The pooled phase scores for each individual were also averaged to determine group mean phase scores for each dependent variable for each phase. The group mean phase scores are reported in Appendix I, are also reported in conjunction with the specific analyses in Chapter 5, and were used as the basis of parametric inferential statistical tests that compared the two groups.

4.2.2.2.2 *Inferential Statistical Tests*

The quasi-experimental design used in this study was chosen in order to avoid the Hawthorne Effect and to extend the data available for assessing the developmental course of SLA. As a result, the design did not include (a) random assignment of learners to the treatment and control groups nor (b) strictly controlled multiple-baselines that would have reduced the length of the PostI phase for some learners. Therefore, this design did not permit the use of any of the various statistical tests designed for specific types of time-series designs (cf. the use of Revusky's *R* in Kennedy, 1988; for further discussion see Mellow, Reeder, & Forster, in press). In

addition, it is important to emphasize that the inferential statistical tests used in this study were not used to generalize to a larger population of language learners, but to the populations of behaviors of these individuals (cf. Mellow, Reeder, & Forster, in press). Two different types of statistical tests were used in this study to compare phase scores across groups, time, task and linguistic context: (a) non-parametric inferential statistical tests, the Friedman Two-way Analysis of Variance, Wilcoxon matched-pairs signed-ranks tests, and Mann Whitney U-tests; and (b) parametric inferential statistical tests, repeated measures ANOVAs.

For the first dependent variable, *the* suppliance for referential definites, non-parametric tests were used to assess group differences and task differences. Non-parametric tests were chosen because of the small sample size in the study which may have resulted in a set of scores that violated several assumptions of parametric tests. These non-parametric statistical tests compared the ranks of the pooled phase scores for each individual on each of the three tasks. To assess task differences for all eight learners during the PreI phase, the Friedman Two-way Analysis of Variance by ranks, a test for matched or related samples, was used. To follow-up this Analysis of Variance test, three two-tailed Wilcoxon matched-pairs signed-ranks tests were used to assess differences between pairs of tests. To assess group differences in gain scores for *the* suppliance for referential definites on each of the three tasks, three Mann Whitney U-tests for two independent samples were performed, one for each task.

For an analysis of group, phase, task, and linguistic context effects in the scores of the four remaining dependent variables, a non-parametric test may have been preferred because of the small sample size in this study. However, a parametric test was selected because the complexity of the data (the number of variables) was greater than could be analyzed by the multivariate non-parametric tests available. The use of these parametric tests may therefore have been problematic because assumptions of normality and homogeneity of variance may have been violated in the scores of this small data set. However, the statistical procedures that were used may be relatively robust to these violations. Tabachnick and Fidell (1989, pp. 378-379) report that MANOVA (and ANOVA) may be robust to violations of normality and homogeneity of variance, especially if the

sample sizes are equal. Because of these limitations regarding the use of these statistical tests, the results should be interpreted with caution.

Three different types of ANOVAs were utilized to analyze group, phase, task, and linguistic context effects in the data. Ideally, an overall MANOVA would have been used first, and then followed-up with additional ANOVAs. However, an overall MANOVA was not possible because of the small number of learners in the study.¹⁸ Therefore, two sets of four repeated measures ANOVAs were performed to analyze group, task, and linguistic context effects. To assess the effects of syntactic position (SPos) and referential distance (RDist) on the spoken and written tasks during the PreI phase for all eight learners, four 2-factor (linguistic context and group (G)) repeated measures ANOVAs were performed (i.e., G x SPos on the spoken task, G x SPos on the written task, G x RDist on the spoken task, G x RDist on the written task). To assess the interaction of syntactic position and referential distance, group, and changes over time (T), four 3-factor doubly repeated measures ANOVAs were performed (i.e., G x T x SPos on the spoken task, G x T x SPos on the written task, G x T x RDist on the spoken task, G x T x RDist on the written task).¹⁹

Because of the small data set available in this study and because of the high variability in the scores, the alpha level for this study was set at 0.1. This level of significance for accepting or rejecting a null hypothesis was chosen to reduce the likelihood of a Type II error, failing to reject

¹⁸ In addition, because *a* overgeneralizations and referentially definite first mentioned NPs were excluded from the data for these four dependent variables, an overall MANOVA using the scores from these variables would have produced slightly different results than the results for the first dependent variable, which provided an overall indication of article use with referential definites.

¹⁹ An additional procedure for assessing the effects and interactions of multiple factors on language use is the VARBRUL statistical procedure (e.g., Cedergren & Sankoff, 1974; for further discussion of the development of VARBRUL, see Fasold, 1990, pp. 249-57). This procedure has recently been used by SLA researchers such as Adamson (1988), Adamson and Regan (1991) and Young (1988, 1989, 1993, 1996) (for further discussion of the use of VARBRUL in SLA research, see Adamson, 1988; Ellis, 1994, pp. 124-25, 151-53; and Preston, 1989, pp. 13-24, 255-60). Although a future VARBRUL analysis may provide valuable additional information about the degrees of relationships between the different contextual variables in this study, the analyses described above were chosen because of their established value for assessing differences in the central tendencies of groups in experimental studies (e.g., Brown, 1988; Hatch & Lartzon, 1991; Seliger & Shohamy, 1989) and because of their widespread public availability in computer programs such as SAS.

the null hypothesis when in fact there is a difference in the scores (cf. Borg & Gall, 1989, p. 351; Brown, 1988, pp. 116-117, 123; Glass & Hopkins, 1984, p. 209-214; Lazarton, 1991, p. 760). In addition, because an overall MANOVA was not used, and instead four ANOVAs were used with each data set, the Bonferroni correction was used to correct the Type I error rate, rejecting the null hypothesis when there is no difference in the scores (cf. Glass & Hopkins, 1984, p. 381; Stevens, 1986, p. 7). Therefore, the overall alpha level of 0.1 was divided by the number of tests (four) to arrive at Bonferroni alpha level of 0.025. Therefore, the alpha level that was used with the ANOVAs in Chapter 5 was 0.025.

4.2.2.2.3 Phase Regression Lines

In order to assess trends in the data, phase regression lines and line graphs of developmental course were utilized. The importance of assessing the trend in the data has already been discussed in sections 1.2, 2.3, 2.5, 3.1 and 3.3.3.3. Although the analyses based on central tendency used the pooled phase score for an individual with respect to each dependent variable, the analyses of trend required a score for each data collection session for each individual. As a result, a score for each individual for each sampling session (or fable version) for each dependent variable was determined using the formulas in 4) through 8) above. As noted above, there was considerable variability in the scores, perhaps due to the small number of contexts, three, that was required to report a session score.

The first type of summary analysis of trend, a phase regression line, was determined for each group for each dependent variable. A regression line provides the best fitting straight line that characterizes the trend (and level) in the scores of a phase. The regression lines for each dependent variable were computed using the session scores of all of the individuals in a group in a phase. Because all of the scores for a group were not independent of each other (i.e., each learner provided more than one score, making those scores dependent), the repeated measurements of the learners were included in the model for the regression lines. The concern about the reliability and consistency of individual session scores in this data set is partially allayed through this aggregation

of all of the session scores for all of the individuals in a group (cf. Cook & Campbell, 1979, p. 43, discussed above). The regression lines are reported graphically and numerically in Appendix J and the numerical values of the slopes are reported in conjunction with the specific analyses in Chapter 5. The comparisons of trends required by the research hypotheses, as discussed in Chapter 5, are particularly concerned with the slope of the lines and relatively less concerned with the y-intercept value.

It was not possible to plot regression lines for each phase for each individual because a minimum of eight data points is needed to plot a regression line (Krishef, 1991, pp. 48-52, 91-100) and only between two and six points were available for the individuals for each phase. The variation in number of points per phase for each individual was due to learners with limited numbers of sampling sessions and due to scores that were not reported because fewer than three contexts of use were present for a dependent variable. The relatively small number of data points had a number of additional implications for the types of analyses used. First, the study did not have enough data points to be a time series experiment (see section 3.3 above), and therefore the differences in trends between phases and groups were not assessed with inferential statistical tests, but only through a simple comparison of slope values. Second, the study did not have enough reliable data points to statistically compare specific details in developmental course (i.e., U-shaped development, delayed effects, etc.); however, these issues are explored in section 5.3 with line graphs. Finally, the use of lines rather than curves has resulted in estimates of trend that have limited accuracy. For example, due to rapid development during the early part of a phase and a relatively small number of data points, the lines sometimes suggested that learners will exceed 100% by the end of the phase (e.g., Appendix J5, J11).

An additional concern regarding the use of regression lines is that they are often used for prediction. For example, given the value for one variable (such as time), the equation for the regression line might allow a researcher to predict the value for another variable (such *the* suppliance). However, regression lines were not used in this study to predict behavior. Because there was considerable variation both between and within group members in *the* suppliance during

the different elicitation sessions, these regression lines accounted for relatively little of the variance in the scores (cf. the small r^2 scores noted in Appendix J).²⁰ Given this variability in these data (i.e., the data points in the scatterplots clearly did not cluster together to form straight lines, see Appendix J), it is important not to have used these regression lines to predict future scores, but simply to have used them as the best available indication of changes in group behavior. In this way, the regression lines were interpreted in much the same way that a mean is interpreted, especially if the scores summarized by a mean have a high standard deviation. As a result, these best fitting lines were used to indicate whether there was a tendency for the scores of the entire group to change during a testing phase. In addition, the regression lines were used to estimate the typical value of a variable during one of the days of the study, specifically days at the beginning and end of a phase. However, this estimation was not a prediction of behavior at times outside of the period of the study. Thus, although regression lines are often used to make inferences (cf. Brown, 1988, p. 133), in the present study regression lines were used in a descriptive manner, describing the trends in the data.

For several reasons, the use of regression lines was particularly effective for assessing the effects of the instructional sessions. First, an examination of trend indicated patterns that were not revealed by mean phase scores (i.e., all post-instruction sessions pooled together) and could not have been indicated by a small number of post-tests. Second, to assess causal relationships, the PostI scores of the treatment group were compared to a set of baseline scores. The use of regression lines allowed comparisons with two different baselines: 1) comparing the level and slope of PostI phase regression lines of the instructed learners to their PreI baseline values (i.e., having the learners act as their own controls, one of the basic premises of the time-series design), to determine if improvement was due to instruction or due to continuing development by the learners; and 2) comparing the changes in slope and level of the scores of the treated learners to the changes in the scores of the control group, assuming that the two groups were roughly equivalent.

²⁰ Indeed, one of the key issues in the study of SLA is to determine why there is so much variation in language use and how researchers can reconcile group behavior and individual behavior (see section 5.3; cf. Mellow & Fuller, 1995).

The between group comparison was of limited value in this study because of the non-equivalence of the groups. In sum, the variability in the scores of these learners suggests that the results reported in Chapter 5 should be interpreted cautiously, but that, in spite of these limitations, the use of regression lines provides much valuable information that would have been lost through the exclusive use of group mean phase scores.

4.2.2.2.4 *Line Graphs*

The second type of summary analysis for trend was the set of lines graphs provided in Appendix B. These line graphs were plotted using the session scores of each individual for each dependent variable. Thus, unlike the three previous summary measures that considered group behavior, this summary analysis considered the individual behavior of the eight learners in the study. Findings from this analysis must be treated with caution because of the relatively small number of measurement sessions per individual (i.e., two to six sessions in a phase) and because of the possible inconsistency in the scores due to the small minimum number of contexts (three) that was required for a session score to be reported. Given these limitations, these line graphs for individuals nonetheless complement the group analyses and provide insight into specific types of responses to instruction, revealing patterns of the developmental course of instructed SLA that include: (a) changes in level that were gradual, immediate, delayed, lasting, and temporary; (b) patterns in trend such as continued improvement and stabilization; and (c) patterns that suggest automatization and restructuring.

In sum, the results reported in Chapter 5 use four types of summary analyses: (a) group mean phase scores, computed using pooled phase scores for each learner; (b) inferential statistical tests, comparing groups and computed using pooled phase scores for each learner; (c) group regression lines, computed using specific session scores for each learner; and (d) line graphs of individuals, using specific session scores for each learner. Group comparisons in sections 5.1 and 5.2 generally use all of the first three analyses. Individual analyses of the developmental course of instructed SLA is considered in section 5.3, using line graphs of individual development.

Chapter 5

Results and Discussion

This chapter reports the results of the eight case studies, attempting to determine when *the* was supplied in a target-like manner by the learners and whether the instruction that four of the learners received affected their suppliance of *the*. The results are reported in three sections, responding to the 12 research hypotheses developed in section 2.5. First, to investigate the first five hypotheses, the scores from the preintervention phase (PreI) data are analyzed, looking at patterns in the data according to task (cloze, composition, speech) and linguistic context (Sub - NSub, Cur - Rec). Each of these variables is considered in terms of the scores of all eight learners and is examined in terms of central tendency and trend. In this section, the equivalence of the groups during the PreI phase is also examined in detail.

The second section of this chapter reports the results of the postintervention phase (PostI), making comparisons with the PreI data both within the treatment group (with the instructed individuals acting as their own controls) and between the two groups. Investigating the effects of instruction, hypotheses 6, 7a, and 7b are addressed. In addition, the second section also investigates the effects of the instruction when analyzed as the interaction between task, linguistic context and instruction, addressing hypotheses 8a, 8b, 9, 10 and 11. This second section considers the dependent variables with respect to central tendency and trend.

The third and final section of this chapter considers the developmental course of instructed second language acquisition, including variations such as immediate or gradual improvement and U-shaped development. In order to assess these properties, the development of individual learners is considered, addressing Hypothesis 12.

5.1. PreI Phase Indications of Interlanguage Patterns

5.1.1 Hypothesis 1: Task Variation, PreI Phase

The predictions for task variation during the PreI phase were developed in section 2.5 and are repeated below.

Hypothesis 1, regarding task variation during the PreI phase: *the* suppliance for referential definites on the cloze task will be more target-like than *the* suppliance on the written task, and *the* suppliance for referential definites on the written task will be more target-like than *the* suppliance on the spoken task, leading to the following hierarchy of scores: cloze task > written task > spoken task.

For the first dependent variable, *the* suppliance for referential definites, the group mean phase scores for each task for all eight learners and for each group are compared in Table 5.1.

Table 5.1
Percentage of the Suppliance for Referential Definites on the Three Different Tasks During the PreI Phase

Task	All learners		Treatment Group		Control Group	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Cloze	94.5	13.0	98.3	3.5	90.8	18.5
Written	60.8	14.9	64.3	7.5	57.3	20.7
Spoken	37.4	20.1	52.8	15.4	22.0	8.8

An examination of these scores confirms the predictions regarding task variation. In order to assess the significance of these task differences, the PreI group mean phase scores for the eight individuals for *the* suppliance for referential definites were considered as a combined group and a Friedman Two-way Analysis of Variance by ranks was performed. The null hypothesis of no differences was rejected: a significant difference for task was found, $Xr^2 = 14.25$, $p < .0001$. To follow up this Analysis of Variance test, three two-tailed Wilcoxon matched-pairs signed-ranks tests were used to assess the ranks of the differences between pairs of tests. These tests revealed that all three tasks were distinct from each other: cloze vs. written ($n=8$, $T=0$) $p = .01$; cloze vs. spoken ($n=8$, $T=0$) $p = .01$; and written vs. spoken ($n=8$, $T=1$) $p = .02$.

This clear task difference for *the* suppliance for referential definites was generally found when the data were sub-divided into specific linguistic contexts, with written suppliance of *the* much higher than spoken suppliance of *the*. These results, and a number of interesting interactions between task and linguistic context are discussed in section 5.1.2 below.

The task differences indicated by this measure of central tendency have a number of important implications. First, these results support the position that greater attention to form leads to more target-like use in interlanguage (cf. Tarone & Parrish, 1988). Second, because the cloze scores were very close to 100%, this indicates that the learners knew the rules for English article use, but were unable to use them as consistently on the spoken and written tasks. This suggests that the learners did have procedural knowledge for using articles in discourse, but that they had not yet automatized this knowledge so that they were able to use it on tasks which do not provide sufficient time (i.e., to make attentional resources available) and focus the learners' attention on article use. The issue of automatization will be discussed in detail below.

In addition to the group mean PreI phase scores, the interlanguage variation on the different tasks can be examined in terms of the rate of development during the PreI phase, as represented by the regression lines (Appendix J). No specific predictions were made, although it was expected that both groups would be developing during this time because the PreI phase occurred during three months of ESL classes in a predominately English-speaking city. The slopes of the regression lines for both groups for all dependent variables on both the spoken and written tasks are provided in Table 5.2.

Table 5.2
Slopes of the Regression Lines for the Spoken and Written Tasks During the PreI Phase for All Five Dependent Variables

	Def		Sub		NSub		Cur		Rec	
	Sp	Wr	Sp	Wr	Sp	Wr	Sp	Wr	Sp	Wr
Treatment	.33	.56	.22	.96	.24	.28	.44	.20	.61	.63
Control	.31	.55	.35	.95	-.22	.21	.35	.84	.22	1.54

An examination of the PreI phase regression lines of the two groups on the different dependent variables reveals the following patterns. The control group was improving much more rapidly on the written task than it was on the spoken task: The rate of development was much more

rapid on the written task for all five variables. This indicates that the task difference for the control group, the Written Task > Spoken Task asymmetry noted above, actually increased during the PreI phase, with the asymmetry at the end of the phase being larger than the asymmetry indicated by the group mean phase scores. In contrast, the treatment group exhibited mixed results with respect to a comparison of trends on the written and spoken tasks: Development on the two tasks was relatively equal in non-subject positions and for recent referents, was faster on the written task in subject position and for referential definites, and was faster on the spoken task for current referents. A further comparison of the development of the two groups is provided in section 5.1.3.

5.1.2 Variation According to Linguistic Context During the PreI Phase

5.1.2.1 Hypothesis 2: Syntactic Position, Spoken Task

The predictions for *the* suppliance on the spoken task in different syntactic positions were developed in section 2.5 and are repeated below.

Hypothesis 2, regarding the effects of syntactic position on the spoken task:
There will be a NSubject > Subject asymmetry in the spoken data (due to efficient omission of *the* in the subject position): *the* will be supplied more often in non-subject positions than in subject position.

The group mean phase scores for *the* suppliance in subject and non-subject positions, for the two groups and on the two tasks, are compared in Table 5.3 below.

Table 5.3
Group Mean Percentage of the Suppliance in Subject and Non-Subject Positions on the Spoken and Written Tasks during the PreI Phase

Task	Treatment Group				Control Group			
	Subject		Non-Subject		Subject		Non-Subject	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Spoken	30.0	22.8	77.5	13.7	5.0	6.2	44.3	15.7
Written	65.8	22.7	69.3	3.3	64.8	30.1	55.0	17.6

An examination of these scores confirms the predictions. On the spoken task, as measured by the group mean phase scores, for both groups *the* was supplied much more for non-subject contexts than for subject contexts (i.e., there were more bare nouns in subject position). In order to assess the significance of these differences, a two factor repeated measures ANOVA was performed on these PreI data. The null hypothesis of no difference was rejected: a significant difference for syntactic position was found, $F = 34.98 (1,6), p < 0.001$. Therefore, as predicted, there was a strong NSub > Sub asymmetry on the spoken task. This asymmetry is consistent with the theoretical position that efficient omission occurs because of limited attentional resources and the redundancy between the two structures, subject (topic) position and use of *the*.

The different syntactic contexts can also be examined in terms of the rate of development during the PreI phase, as represented by the slopes of the regression lines, which have already been provided in Table 5.2. No specific predictions were made, although it was expected that both groups would be developing during this time. The relatively small values of the slopes on the regression lines (less than .40, with one negative value) on the spoken task indicate that both groups exhibited only moderate changes during the PreI phase for both syntactic positions. The slopes reveal that the treatment group exhibited no difference in PreI phase rate of improvement for subject vs. non-subject position (.22 vs. .24, respectively), but that the control group did (.35 vs. -.22). This difference, with improvement in subject position and a decrease in non-subject positions, suggests that the NSub > Sub asymmetry decreased for the control group during the PreI phase, and reversed to a Sub > NSub asymmetry by the end of the PreI phase (Day 85 predictions, based on the regression lines: Sub: 20.2; NSub: 12.4). However, as discussed in section 5.2.3.1 below, this PreI phase development may have resulted from some unusual fluctuations in learner behavior.

5.1.2.2 Hypothesis 3: Syntactic Position, Written Task

The predictions for *the* suppliance in different syntactic position on the written task were developed in section 2.5 and are repeated below.

Hypothesis 3, regarding the effects of syntactic position on the written task:
 There will be a Sub > NSub asymmetry in the written data (due to priming of *the* in subject position): *the* will be supplied more often in subject position than in non-subject positions. However, these effects may be moderated or obscured by an additional factor: The application of metalinguistic knowledge may affect *the* suppliance on the written task.

An examination of the scores in Table 5.3 does not confirm the predictions regarding an asymmetry with regard to syntactic position. In the written task, as measured by the group mean phase scores, for the control group *the* was supplied somewhat more for subject contexts than for non-subject contexts (i.e., there were more bare nouns in non-subject position) and for the treatment group *the* was supplied almost equally in the two syntactic positions. In order to assess the significance of these differences, a two factor repeated measures ANOVA was performed on these PreI data. The null hypothesis of no difference was not rejected: a significant difference for syntactic position did not emerge, $F = 0.16 (1,6), p = .70$ (n.s.). On the basis of the measure of central tendency, the group mean phase scores, the prediction of priming asymmetry was not supported.

However, an analysis of the regression lines for the written data indicates that the use of a group mean phase scores for the phase is actually quite misleading for these data. As can be seen in the slope of the regression lines in Table 5.2 above and in the graphs in Appendix J, there was a large difference in the PreI phase slopes of the two syntactic contexts. The rate of development for *the* suppliance in subject position was much higher than the rate of development for use in non-subject position. For the treatment group, the slope for subject position was .96, whereas the slope for non-subject position was .28. Thus, when a regression line is used to summarize the trend in the phase for the treatment group, the equation of that line indicates that subject position group scores improved by 77 percentage points during the PreI phase (using a predicted score at 80 days, to provide comparable values for both groups). In contrast, the equation for non-subject position for the treatment group suggests that those scores increased by only 22 percentage points during the phase.

For the control group, the slope for subject position was .95, whereas the slope for non-subject position was only .21. The equation for subject position indicates that the group scores would have improved by 76 percentage points during the 80 day period. In contrast, the equation for non-subject position suggests an improvement of only 17 percentage points during the PreI phase. These predicted differences in scores at the beginning and end of the PreI phase are illustrated in Table 5.4.

Table 5.4
Predicted Scores for Percentage of the Suppliance on Day 1 and Day 80 in Subject and Non-Subject Positions on the Written Task During the PreI Phase

Treatment	Day 1		Day 80	
	Subject	Non-Subject	Subject	Non-Subject
Treatment	44	59	121 ^a	81
Control	29	27	105 ^a	44

^a These numbers are greater than 100% and indicate that a ceiling, of perhaps 100%, would have been reached before the end of the phase.

These differences in slope indicate that at the very beginning of the study there was a NSub > Sub asymmetry in the scores of the treatment group, that is, the asymmetry expected if attentional resources were limited and efficiency was a dominant factor. The predicted day 1 scores for the treatment group are 44% (subject position) and 59% (non-subject positions). The predicted day 1 scores indicate that the control group supplied *the* relatively equally in both syntactic positions. The differences in slope also indicate that at the end of the PreI phase there was a clear Sub > NSub asymmetry in the scores of both groups, that is, the asymmetry expected if attentional resources were relatively abundant and priming was a dominant factor. As indicated in Table 5.4, the predicted end of phase scores for the treatment group are 100% in the subject position and 81% in the non-subject positions. The predicted end of phase scores for the control group are 100% in the subject position and 44% in the non-subject position.

Although the differences between these day 80 predicted scores cannot be subjected to inferential statistical tests, and although there is considerable variation in the data sets, this

longitudinal data is able to reveal important and dynamic properties of the interlanguage systems of these learners. The regression lines indicate that the prediction of a priming asymmetry (Sub > NSub) was valid for both groups on the written task, but only for language use near the end of the phase. However, what remains unaccounted for is the early NSub > Sub asymmetry of the treatment group and the subsequent development of both groups. The moderating factor discussed in Principle 5, application of metalinguistic knowledge, would not account for this interesting developmental trend because the asymmetries do not appear to have been moderated, nullified or obscured by any conscious rule application. For conscious knowledge to result in these patterns, the learners would have needed to gradually learn and use a conscious rule that favoured article provision only in subject position. Perhaps, the most plausible explanation is that the learners found the written task to be difficult at the beginning of the data collection period and therefore their attentional resources were fully taxed. However, as the learners became more proficient at the task, their attentional resources were no longer fully taxed and priming led to the observed asymmetry. In sum, part of these complex data appear to be explained by the Redundancy Principle, hence providing support for principles developed in section 2.4.

5.1.2.3 Hypothesis 4: Referential Distance, Spoken Task

The predictions for *the* suppliance for different referential distances on the spoken task were developed in section 2.5 and are repeated below.

Hypothesis 4, regarding the effects of referential distance on the spoken task:
There will be a Rec > Cur asymmetry in the spoken data: *the* will be supplied more often for recent NPs than for current NPs.

The group mean phase scores for *the* suppliance with current and recent NPs, for the two groups and on the two tasks, are compared in Table 5.5 below.

Table 5.5
Percentage of the Suppliance for Current and Recent NPs on the Spoken and Written Tasks During PreI Phase

Task	Treatment Group				Control Group			
	Current		Recent		Current		Recent	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Spoken	45.5	17.7	56.5	16.1	17.2	6.4	24.0	14.0
Written	67.8	14.3	68.5	13.0	59.5	24.5	63.5	25.5

An examination of these scores confirms the predictions regarding referential distance. On the spoken task, as measured by the group mean phase scores, for both groups *the* was supplied more for recent referents than for current referents (i.e., there were more bare nouns used for current NPs). In order to assess the significance of these differences, a two factor repeated measures ANOVA was performed on these PreI data. The null hypothesis of no difference was rejected: a significant difference for referential distance was found, $F = 10.388 (1,6), p < 0.02$. Therefore, as predicted, there was a Rec > Cur asymmetry on the spoken task.

The recent and current referents can also be examined in terms of the rate of development during the PreI phase, as represented by the slopes of the regression lines, which have already been provided in Table 5.2. No specific predictions were made, although it was expected that both groups would be developing during this time. The slopes indicate that there were differences in the rate of development for these linguistic contexts on the spoken task: Treatment: Current .44 vs. Recent .61, and Control: Current .35 vs. Recent .22. All learners therefore improved in their provision of *the*, with the Rec > Cur asymmetry increasing for the treatment group (to a difference of 20.6% by day 80), and the Rec > Cur asymmetry disappearing for the control group (to a Cur > Rec asymmetry of 4.4% by day 80).

Although the Rec > Cur asymmetry was not as large as the NSub > Sub asymmetry, and although the asymmetry appears to have disappeared by the end of the PreI phase for the control group, the Rec > Cur asymmetry in these spoken data is consistent with the theoretical position

that a polyfunctional form will be supplied more often than a unifunctional or counterfunctional form. In other words, a form was predicted to be supplied more often in a context where two or more functions are fulfilled by its use, in comparison to the supplience of that form in a context where only one function would be fulfilled by its use and another function will be nullified or reversed. Thus, this trend is consistent with and supportive of the polyfunctional principle (P2) and Givon's (1983) notions of topic continuity and the quantity universal.

5.1.2.4 Hypothesis 5: Referential Distance, Written Task

The predictions for *the* supplience for different referential distances on the written task were developed in section 2.5 and are repeated below.

Hypothesis 5, regarding the effects of referential distance on the written task:
There will be a Rec > Cur asymmetry in the spoken data: *the* will be supplied more often for recent NPs than for current NPs. However, these effects may be moderated or obscured by an additional factor: The application of metalinguistic knowledge may affect *the* supplience on the written task.

An examination of the scores in Table 5.5 does not confirm the predictions regarding an asymmetry with respect to referential distance. In the written task, as measured by the group mean phase scores, for both groups *the* was supplied relatively equally for both recent contexts and current contexts, with a very slight trend to supply *the* more often for recent referents than for current referents. In order to assess the significance of these differences, a two factor repeated measures ANOVA was performed on these PreI data. The null hypothesis of no difference was not rejected: $F = 0.17 (1,6)$, $p = .69$ (n.s.).

The current and recent referents can also be examined in terms of the rate of development during the PreI phase, as represented by the slopes of the regression lines, which have already been provided in Table 5.2. No specific predictions were made, although it was expected that both groups would be developing during this time. The slopes for the written data indicate that recent NPs were developing much faster than current NPs: Treatment: Current .20 vs. Recent .63, and Control: Current .84 vs. Recent 1.54, suggesting that a Rec > Cur asymmetry appeared during the

second half of the phase. In addition, the regression lines indicate that both groups reached a ceiling (very close to 100%) before the end of the PreI phase.

The absence of a strong Rec > Cur asymmetry does not invalidate Hypothesis 5 above. Although it was hypothesized that polyfunctionality may lead to an asymmetry, it was also hypothesized that the effects of referential distance could be nullified or obscured on the written task by the conscious use of grammatical knowledge. This seems to be a plausible explanation because *the* suppliance was much higher on the written task than on the spoken task, as indicated in Table 5.5. Thus, it seems reasonable that the use of grammatical knowledge obscured the effects of referential distance. Although it is possible that referential distance is not a factor in the written interlanguage behavior of these learners, another reason why the Rec > Cur asymmetry was absent in the written data could be the manner in which referential distance was operationalized. Although the method used in this study, counting the number of clauses since previous mention, is practical and reliable, it may not be as valid as other methods of determining referential distance, such as considering the role and length of pauses or asking learners about the assumptions that they are making about the knowledge of listeners. In sum, effects of referential distance were not found on the written task. Because an effect for referential distance was found on the spoken task, these data are consistent with Principle 5, that the spoken task will more clearly exhibit systematic interlanguage patterns than will the written task.

Overall, this examination of article use in two linguistic contexts during the PreI phase has provided three important conclusions. First, the use of a measure of central tendency (or perhaps any static measure) to represent pre-test language ability is clearly very limited in its ability to represent the actual pre-intervention abilities of the learners. This was most clearly seen in the changes during the PreI phase in the learners' interlanguage system involving the provision of articles in subject and non-subject positions on the written task. This issue is considered in greater detail in section 5.2. Second, although both linguistic contexts account for some variation in the provision of *the*, there appears to be more systematic variation in the suppliance of *the* due to syntactic position and somewhat less systematic variation in the suppliance of *the* in response to

referential distance. Third, if the theoretical accounts provided are correct, it appears that both the efficient reduction of redundancy and the reinforcement provided by polyfunctionality are processes that largely appear on tasks that make considerable demands on limited attentional resources, including the spoken fable retelling and possibly including the first few written tasks. However, when tasks do not make large demands on attentional resources, such as written tasks, then grammatical knowledge may be used, drowning out the effects of factors such as referential distance. In addition, when tasks do not make large demands on attentional resources, especially practiced compositions, then priming may have a strong effect, an effect that may add to conscious grammatical knowledge rather than being obscured by such metalinguistic knowledge.

5.1.3 *An Examination of the Equivalence of the Groups*

Although the eight volunteer learners were fairly homogeneous in certain respects, the two groups were not equivalent (cf. 4.1.1). As indicated in the PreI phase data, many aspects of which were discussed in the preceding section, the two groups were similar in many respects. First, on the cloze and written tasks, the control and treatment groups were relatively comparable: The treatment group was slightly better on the cloze task, 98% > 91%, and, on the compositions, 64% > 57%, for the first dependent variable, *the* suppliance for referential definites. The differences between the groups on the written task were subjected to two, 2-factor repeated measures ANOVAs, one that considered syntactic position and one that considered referential distance. The null hypotheses of no difference between groups were not rejected, that is, the groups performed similarly on the written task (syntactic position: $F = 0.37 (1,6)$, $p = 0.56$ (n.s.); referential distance: $F = 0.26 (1,6)$, $p = 0.63$ (n.s.)).

Second, an examination of *the* suppliance for referential definites revealed that both groups exhibited the same ranking of task accuracy in the PreI phase (Treatment: Cl = 98.3 > Wr = 64.3 > Sp = 52.8; Control Cl = 90.8 > Wr = 57.3 > Sp = 22; as discussed in section 5.1.1). Third, the regression lines indicated that the control group and treatment group developed at similar rates

during the PreI phase for six of the ten dependent variables. The slopes of the regression lines for those variables, a subset of the values in Table 5.2, are listed in Table 5.6 below (cf. Appendix J).

Table 5.6
Slopes of the Regression Lines for Treatment and Control Groups, Spoken (Sp) and Written (Wr) Tasks, PreI Phase, for Six Dependent Variables

Group	Dependent Variable					
	Sp Def	Sp Sub	Sp Cur	Wr Def	Wr Sub	Wr NSub
Treatment	.33	.22	.44	.56	.96	.28
Control	.31	.35	.35	.55	.95	.21

Thus, for these six dependent variables, the development of *the* suppliance during the PreI phase was similar for both groups on the spoken task.

The relations between *the* provision in the complementary linguistic contexts was also similar for the two groups. Thus, a fourth and very strong similarity between the treatment and control groups was that for the group mean phase scores of *the* suppliance, the same relation between the two sets of complementary linguistic contexts existed on both the spoken and written tasks, as illustrated in Table 5.7 below.

Table 5.7
Patterns in Percentage of the Suppliance for Treatment and Control Groups According to Linguistic Context, PreI Phase, Assessed with Group Mean Phase Scores

Group	Patterns According to Task and Linguistic Context							
	Sp NSub > Sub		Wr NSub ≈ Sub		Sp Rec > Cur		Wr Rec ≥ Cur	
Treatment	77.5	30	69.3	65.8	56.5	45.5	68.5	67.8
Control	44.3	5	55	64.8	24	17.2	63.5	59.5

As Table 5.7 reveals, both groups exhibited a large NSub > Sub asymmetry on the spoken task, a moderate Rec > Cur asymmetry on the spoken task, a very slight Rec > Cur asymmetry on the written task, and relatively comparable scores in both syntactic positions on the written task.

A fifth similarity between the treatment and the control groups is that, for the slopes of the regression lines (i.e., the rates of development) during the PreI phase, the relations between the

sets of complementary linguistic contexts were the same for the written task, as illustrated in Table 5.8.

Table 5.8
Patterns in Slope Values of the Suppliance on the Written Task for Treatment and Control Groups According to Linguistic Context, PreI Phase

Group	Patterns According to Task and Linguistic Context			
	Wr Sub > NSub		Wr Rec > Cur	
Treatment	.96	.28	.63	.20
Control	.95	.21	1.54	.84

As Table 5.8 reveals, on the written task both groups exhibited faster development in subject position and faster development for recent referents. As noted above, both of these developmental trends support the predictions in the research hypotheses, with the Sub > NSub slope values leading to anticipated priming effects and with the Rec > Cur slope values leading to the anticipated polyfunctionality effects.

Although the data exhibited these five similarities, the two groups were different in a number of respects, many aspects of which were discussed in the preceding section. First, and perhaps most important, on the spoken task, the groups were clearly non-equivalent. For example, for *the* suppliance for referential definites, the treatment group averaged 53% and the control group averaged only 22%. The differences between the groups on the spoken task were subjected to two, 2-factor repeated measures ANOVAs, one that considered syntactic position and one that considered referential distance. The null hypotheses of no difference between groups were rejected, that is, the groups performed differently on the spoken task (for syntactic position: $F = 12.04$ (1,6), $p = 0.01$; for referential distance: $F = 9.82$ (1,6), $p = 0.02$)).

A second difference between the two groups is that on four of the ten dependent variables the control group was developing at a different rate than the treatment group. The slopes of the regression lines for those four dependent variables are listed in Table 5.9 below (cf. Appendix J).

Table 5.9
Slopes of the Regression Lines for Treatment and Control Groups, Spoken (Sp) and Written (Wr) Tasks, PreI Phase, for Four Dependent Variables

Group	Dependent Variable			
	SP NSub	Sp Rec	Wr Cur	Wr Rec
Treatment	.24	.61	.20	.63
Control	-.22	.22	.84	1.54

For these variables, the treatment group was developing faster than the control group on the spoken task, but slower than the control group on the written task.

A third difference between the two groups is that for the slopes of the regression lines (i.e., the rates of development) during the PreI phase, the relations between the sets of complementary linguistic contexts were different on the spoken task, as illustrated in Table 5.10.

Table 5.10
Patterns in Slope Values of the Suppliance on the Spoken Task for Treatment and Control Groups According to Linguistic Context, PreI Phase

Group	Patterns According to Task and Linguistic Context			
	Sp NSub	Sub	Sp Rec	Cur
Treatment	.24	≈ .22	.61	> .44
Control	-.22	< .35	.22	< .35

As Table 5.10 reveals, the treatment group exhibited similar development on both syntactic positions on the spoken task, but the control group exhibited faster development in subject position. In addition, whereas the treatment group developed faster for recent referents than for current referents, the control group exhibited the opposite pattern, developing faster for current referents than for recent referents.

Overall, the two groups are similar in many respects, but clearly are different in other aspects of their suppliance of *the*. This overall non-equivalence threatens the validity of between group comparisons, and indicates the need for the learners to act as their own controls.

5.2. The Effects of Instruction on the Five Dependent Variables

Having determined, in detail, the nature of the baseline behavior of these eight learners, and having found that the data appear to be explained by several theoretical principles, the analyses in this section consider the main question of the dissertation: What are the effects of the instruction? This question is answered in the following manner. First, using two different measures of group behavior during each phase, each of the five dependent variables is considered separately. In section 5.2.1, each variable is assessed through an examination of the gains in the group mean phase scores, with the differences between the treatment and control groups assessed with inferential statistical tests. Given the limitations of this use of means, a measure of central tendency, section 5.2.2 presents an assessment of the five variables through an examination of the differences between PreI phase and PostI phase group regression lines, making comparisons within the treated group and between the two groups. After these two methods are used to assess each dependent variable, the variables are then considered in relation to each other. Therefore, in section 5.2.3, the patterns in the changes in interlanguage systems are considered, that is, changes in the relations (e.g., asymmetries) between the complementary linguistic contexts, syntactic position and referential distance. These potential changes in interlanguage systems are examined through a consideration of the gains in the group mean phase scores and through a consideration of potential differences in PostI phase regression lines, both within groups and between groups.

5.2.1 Hypothesis 6: Group Differences, Group Mean Phase Scores, All Dependent Variables, All Tasks

The predictions for effects of instruction on all five dependent variables on all tasks, assessed with group mean phase scores, were developed in section 2.5 and are repeated below.

Hypothesis 6, regarding group differences in gains in mean phase scores:
With respect to the five dependent variables on the written, spoken, and cloze tasks, the instructed group will exhibit greater gains in mean phase scores from the PreI phase to the PostI phase than will the uninstructed group.

Gain scores, as group mean phase scores for *the* suppliance, for all five variables for the two groups and on the two tasks (a third task for all definite contexts), are compared in Table 5.11 below (a full chart of individual mean phase scores and group means is provided in Appendix I).

Table 5.11
Gain Scores, Based on the Group Mean Percentages of the Suppliance on the Three Different Tasks for Each of Five Dependent Variables

		Treatment Group									
		Def		Sub		NSub		Cur		Rec	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Spoken	pre-	52.8	15.4	30.0	22.8	77.5	13.7	45.5	17.7	56.5	16.1
	post-	79.8	16.4	74.3	26.9	84.8	90.0	79.8	20.7	79.3	15.3
	gain	27.0		44.3		7.3		34.3		22.8	
Written	pre-	64.3	7.5	65.8	22.7	69.3	3.3	67.8	14.3	68.5	13.0
	post-	89.5	12.5	93.3	11.6	90.8	9.3	92.3	13.0	94.5	7.1
	gain	25.3		27.5		21.5		24.5		26.0	
Cloze	pre-	98.3	3.5								
	post-	91.3	13.3								
	gain	-7.0									
		Control Group									
		Def		Sub		NSub		Cur		Rec	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Spoken	pre-	22.0	8.8	5.0	6.2	44.3	15.7	17.3	6.4	24.0	14.0
	post-	42.3	27.9	41.3	28.0	47.8	29.6	38.8	32.6	48.5	26.3
	gain	20.3		36.3		3.5		21.5		24.5	
Written	pre-	57.3	20.7	64.8	30.1	55.0	17.6	59.5	24.5	63.5	25.5
	post-	60.5	22.6	81.8	15.9	51.5	25.3	71.8	17.8	59.8	27.5
	gain	3.3		18.5		-3.5		12.3		-3.8	
Cloze	pre-	90.8	18.5								
	post-	83.5	21.1								
	gain	-7.3									

The data from the cloze task indicate that the learners were close to 100% in *the* suppliance for referential definites during both the PreI and PostI phases. As noted above in section 5.1.1, if the cloze task focuses the learners' attention on article use and allows them to make use of their grammatical knowledge, then these results indicate that these learners have non-automatized procedural knowledge for using *the* in discourse. The data indicate that *the* suppliance for

referential definites decreased slightly for both groups during the PostI phase. This slight decrease is most probably due to a lack of reliability in the scores: a small number of instances of non-suppliance of *the* during the PostI phase (five for the treatment group, up from one during the PreI; and four for the control group, up from three during the PreI phase) led to this variation because of the small number of contexts (three) per cloze task. Therefore, these data suggest that throughout the study, the learners in both groups maintained a target-like ceiling with respect to the suppliance for referential definites on the cloze task. Consequently, the results for the cloze task will not be considered further in this section.

The research hypothesis predicted that there would be greater gains for the treatment group than for the control group. These predictions appear to be confirmed in these data, with the relevant differences in gain scores (derived from Table 5.11) presented in Table 5.12.

Table 5.12
Differences between Group Gains in Mean Percentage of the Suppliance (i.e., TrGgain - ConGgain) on the Spoken and Written Tasks for Each of the Five Dependent Variables

	Def	Sub	NSub	Cur	Rec
Spoken	6.7	8	3.8	12.8	-1.7
Written	22	9	25	12.2	29.8

For nine out of ten comparisons on the spoken and written tasks, the treatment group improved more than the control group, with the greater gain for the treatment group ranging from 3.8 percentage points to 29.8 percentage points. Only for recent NPs (Rec) on the spoken task did the control group improve more than the treatment group (gains of 24.5% vs. 22.8%, respectively), with that difference being so slight (-1.7) as to suggest that both groups improved equally. In addition, the advantage for the treatment group was greater on the written task than on the spoken task.

In order to assess the significance of the group differences for *the* suppliance for referential definites (Def) for each task, three non-parametric Mann-Whitney *U* tests were performed on the scores. First, in order to assess the significance of the difference between the gains of the treatment and control groups on the spoken task for *the* suppliance for referential definites, a one-

tailed Mann-Whitney U test was performed on the PreI and PostI pooled phase scores of each individual (the group means and standard deviations are reported in Table 5.11 above). The null hypothesis of no difference was not rejected: a significant difference for group by phase was not found, $U = 5.5$ ($n_1 = 4$, $n_2 = 4$), $p = .34$ (n.s.).

Second, in order to assess the significance of the difference between the gains of the treatment and control groups on the written task for *the* suppliance for referential definites, a one-tailed Mann-Whitney U test was performed on the PreI and PostI pooled phase scores of each individual (the group means and standard deviations are reported in Table 5.11 above). The null hypothesis of no difference was not rejected: a significant difference for group by phase was not found, $U = 4$ ($n_1 = 4$, $n_2 = 4$), $p = .17$ (n.s.).

Third, in order to assess the significance of the very slight difference between the gains of the treatment and control groups on the cloze task for *the* suppliance for referential definites, a one-tailed Mann-Whitney U test was performed on the PreI and PostI pooled phase scores of each individual (the group means and standard deviations are reported in Table 5.11 above). The null hypothesis of no difference was not rejected: a significant difference for group by phase was not found, $U = 7$ ($n_1 = 4$, $n_2 = 4$), $p = .44$ (n.s.). Thus, the differences in the group mean gain scores for these three variables were not statistically significant.

In order to assess the significance of the group differences for the remaining four variables on the written and spoken tasks, four doubly repeated ANOVAs were performed on the scores (cf. section 4.2.2.2 for a discussion of the selection of these parametric statistical tests). First, in order to assess the significance of the differences between the gains of the treatment and control groups on the spoken task for *the* suppliance in subject and non-subject syntactic positions, a three factor doubly repeated measures ANOVA was performed on the PreI and PostI phase pooled scores of each individual (the group means and standard deviations are reported in Table 5.11 above). The null hypothesis of no differences was not rejected: a significant difference for group by phase was not found, $F = 0.23$ (1, 6), $p = .64$ (n.s.).

Second, in order to assess the significance of the differences between the gains of the treatment and control groups on the spoken task for *the* suppliance with current and recent NPs, a three factor doubly repeated measures ANOVA was performed on the PreI and PostI phase pooled scores of each individual (the group means and standard deviations are reported in Table 5.11 above). The null hypothesis of no differences was not rejected: a significant difference for group by phase was not found, $F = 0.18 (1, 6), p = .69$ (n.s.).

Third, in order to assess the significance of the differences between the gains of the treatment and control groups on the written task for *the* suppliance in subject and non-subject syntactic positions, a three factor doubly repeated measures ANOVA was performed on the PreI and PostI phase pooled scores of each individual (the group means and standard deviations are reported in Table 5.11 above). The null hypothesis of no differences was not rejected: a significant difference for group by phase was not found, $F = 1.85 (1, 6), p = .22$ (n.s.).

Fourth, in order to assess the significance of the differences between the gains of the treatment and control groups on the written task for *the* suppliance with current and recent NPs, a three factor doubly repeated measures ANOVA was performed on the PreI and PostI phase pooled scores of each individual (the group means and standard deviations are reported in Table 5.11 above). The null hypothesis of no differences was not rejected: a significant difference for group by phase was not found, $F = 2.10 (1, 6), p = .20$ (n.s.). In sum, the differences in the group mean gain scores for all eleven variables, including the eight that were assessed with ANOVAs, were not statistically significant.

Although the descriptive statistics reported in Tables 5.10 and 5.11 indicated that the treatment group improved more than the control group for nine of the eleven variables, these differences were not found to be statistically significant: the null hypotheses of no differences were not rejected. A preliminary interpretation of these analyses might be that the instruction did not have a *meaningful* effect on the language use of these learners. However, such an interpretation may be spurious for a number of important reasons. First, a careful examination of the results of the individuals in the study (cf. section 5.3 and Appendix B) reveals that all four instructed

learners improved considerably after instruction on both written and spoken tasks, that three of uninstructed learners exhibited no change or became less target-like during the PostI phase, and that one of the uninstructed learners improved considerably during the PostI phase. With such a small sample (four per group), the improvement by the one uninstructed learner prevented the difference between the treatment and control groups from achieving a statistically significant level on the statistical tests. However, for one learner to improve in *the* suppliance without specific instruction is not particularly surprising because the learners were all living in an English speaking context, sharing apartments with English speakers, and taking pre-university courses in English (cf. section 4.1.1). As a result, a conclusion for these learners would be that tutored form-focused instruction was always followed by improvements, and that only one out of four learners improved without instruction. Determining that instruction is always (or even often) effective, while the extensive use of a second language only occasionally results in comparable development, would seem to be a pedagogically meaningful result, even if it is not statistically significant. Therefore if these findings were found to be generalizable beyond these learners and this language item, through future replications, then these results would seem to be educationally relevant.

A second concern about the interpretation of these statistical tests arises because the groups are non-equivalent, as indicated by the self-selected nature of the groups and by the examination of the PreI phase scores (section 5.1.3). This non-equivalence could mean that the self-selected treated group was more motivated in general and improved more because they benefited more from their English experiences inside and outside of the classroom, but not because of the instruction. However, this non-equivalence could also mean that, because the control group generally started at a lower ability level, they had more room to improve. In contrast, the treatment group started at a higher ability level, closer to the ceiling, and may have made very meaningful progress even though that progress was not exhibited in statistically significant gain scores. This second interpretation of non-equivalence seems to be particularly valid because the one control learner who made considerable progress during the PostI phase began the study at a particularly low level. As indicated in Appendix B, for *the* suppliance for referential definites, the third control learner

(Con3) had PreI phase mean scores lower than any of the treated learners on both the written and spoken tasks. As a result, Con3's gain scores were greater than any treated learner, but Con3's PostI phase mean scores were higher than only either one or two of the treated learners, on the written and spoken tasks, respectively. Because the statistical tests that were used compared gain scores, the non-equivalence during the PreI phase contributed to the non-significant result. If PostI phase scores, rather gain scores, of *the* suppliance for referential definites were compared for the two groups using the Mann-Whitney *U* tests, then the *p* values for both the spoken and written tasks would have been .057 (cf. the discussion of Doughty, 1991 in section 3.2). Although such a comparison of only PostI phase scores suffers from a number of threats to internal validity, the difference in results between the use of gain scores compared to the use of PostI phase scores illustrates the importance of determining whether the use of specific measures leads to results that are pedagogically meaningful. In addition, the concern about the non-equivalence of groups (already discussed in section 5.1.3), combined with the very small sample size in this study, re-emphasizes the concerns about internal validity used in this type of study. Therefore, in the analyses in the next two sections the learners also act as their own controls in order to increase confidence in the attribution of causality.

Understanding and interpreting the non-equivalence of groups is related to a third concern about the interpretation of these statistical tests: the validity of interval scales for measuring interlanguage development (cf. section 3.3.3.3). In other words, there may have been a significant difference between the two groups, but the scale that was used was not able to measure it. This may be particularly true due to ceiling effects (cf. Principle 6 in section 2.4). For example, the interval scale that was used for this study assumes the gains made from 39% to 79% (a 40% gain) are greater than gains made from 59% to 94% (only a 35% gain). These were the gains made, respectively, by the third control learner (Con3) and the third treated learner (Tr3) in *the* suppliance for referential definites on the written task (see Appendix B). However, given that achieving 100% target-like use may be particularly difficult because of ceiling effects, it might be argued that Tr3, by progressing to 94%, made greater progress than Con3, who progressed to

79%. Thus, the construct validity of measurement systems must be carefully considered. Lantolf and Frawley (1988) referred to this problem as the psychometric posture: the assumption that behavior is actually interval or metric simply because the measurement scale gives it those values. Mellow and Fuller (1995) provided an analysis of the development of these learners that does not adopt these psychometric assumptions.

A fourth concern about the interpretation of these statistical tests, one that is addressed in the next section, is an additional concern about the validity of measurement. In the analyses noted above, a measure of central tendency was used: the means of each groups' pooled phase score. As noted above in section 5.1.2.2, a static measure such as a mean score may not reflect changes in the developing systems of language learners: a mean or a single data point for a phase may not indicate that the learner was improving, or that the learner had stabilized.²¹ This fourth concern suggests that an examination of group regression lines is able to provide additional insight into the effect of instruction on these language learners, including comparisons within groups as well as between groups.

5.2.2 PreI - PostI Changes in the Regression Lines

Before addressing hypotheses 7a and 7b, this section summarizes the regression graphs in Appendix J, revealing a number of important characteristics of the development of the learners in this study. Table 5.13 summarizes certain aspects of the regression lines provided in Appendix J. For each variable, Table 5.13 provides the estimated scores for the first day of the PostI phase (day 92 (D92) for the treatment group and day 85 (D85) for the control group, due to slight differences in scheduling), based on both the PreI phase group regression line (preRL) and the PostI phase group regression line (postRL). For each variable, Table 5.13 also provides the slopes from the group regression line for each phase. Finally, Table 5.13 also provides the change

²¹ Anticipated stabilization, as indicated by Mellow, Reeder, and Forster (in press), can only be verified by the use of an equivalent control group (i.e., equivalent in level and in learner behavior, in the likelihood of studying and practicing, and in the likelihood of using specific strategies on the elicitation tasks).

in level estimated for the first day of the PostI phase, as well as the change in slope between the two phases.

Table 5.13
Slope Values and Estimated Scores on the First Day of the PostI Phase of Percentage of the Suppliance for the Treatment and Control Groups on the Spoken and Written Tasks for Each of Five Dependent Variables

		Treatment Group									
		Def		Sub		NSub		Cur		Rec	
		D92	Slope	D92	Slope	D92	Slope	D92	Slope	D92	Slope
Spoken	preRL	69.7	.33	46.1	.22	91.9	.24	72.1	.44	94.7	.61
	postRL	73.6	.41	72.5	.50	71.4	.40	69.6	.55	81.7	.23
	change	3.9	.08	26.4	.28	-20.5	.16	-2.5	.11	-13.0	-.38
Written	preRL	92.8	.56	132.4	.96	84.7	.28	97.5	.20	119.5	.63
	postRL	88.1	.12	95.0	-.03	84.9	.32	90.6	.08	102.4	-.05
	change	-4.7	-.44	-5.0 ^a	-.99	0.2	.04	-6.9	-.12	0.0 ^a	-.68
Cloze	preRL	100	.00								
	postRL	99.9	.01								
	change	-0.1	.01								
		Control Group									
		Def		Sub		NSub		Cur		Rec	
		D85	Slope	D85	Slope	D85	Slope	D85	Slope	D85	Slope
Spoken	preRL	27.6	.31	20.2	.35	12.4	-.22	26.5	.35	21.4	.22
	postRL	27.6	.08	28.7	-.04	26.9	.04	19.3	.22	39.8	-.10
	change	---	-.23	8.5	-.39	14.5	.26	-7.2	-.13	18.4	-.32
Written	preRL	75.1	.55	109.8	.95	44.5	.21	108.0	.84	119.0	1.54
	postRL	50.1	.01	88.5	-.29	30.8	.13	55.0	-.01	62.4	.03
	change	-25.0	-.54	-11.5 ^a	-1.24	-13.7	-.08	-45.0 ^a	-.85	-37.6 ^a	-1.51
Cloze	preRL	111.3	.27								
	postRL	73.4	.13								
	change	-26.6 ^a	-.14								

^a These change scores were calculated after the predicted values had been adjusted to 100%.

The analyses in this section compare 20 sets of regression lines: The changes from the PreI phase to the PostI phase for each of the five dependent variables on the spoken and written tasks for each group. Each of these twenty changes is characterized as belonging to one of the following four categories: 1) gains in target-like use, 2) continued development of target-like use or

maintenance of a very high target-like use, 3) continued stabilization at a level below a ceiling, and 4) losses in target-like use. These categories involve changes between the two phases in terms of both level and slope. Due to the limited accuracy of regression lines based upon a relatively small number of data points, concerns about the nature of the measurement scale, and the variability of learner suppliance of *the*, changes in level of less than 10% are not considered to be a change in level and changes in slope of less than .15 are not considered to be a change in level. These choices of cut-off levels had the following results. The following were considered to be changes in level: -11.5, -13.0, -13.7, 14.5, 18.4, -20.5, -25.0, 26.4, -37.6, and -45.0. The following were considered to not be changes in level: 8.5, -7.2, -6.9, -5.0, -4.7, 3.9, -2.5, 0.2, 0.0 and 0.0. The following were considered to be changes in slope: .16, -.23, .26, .28, -.32, -.38, -.39, -.44, -.54, -.68, -.85, -.99, -1.24, and -1.51. The following were considered to not be changes in slope: -.13, -.12, .11, -.08, .08, and .04. A summary of the categorizations is provided in Table 5.14.

Table 5.14
Categories of Changes in Regression Lines from the PreI Phase to the PostI Phase

-
1. Gains in target-like use
 1. Gains in level and slope: TrG Sub Sp
 2. Continued development or maintenance of target-like use
 - 2a. No change in level, continued positive slope: TrG Def Sp; TrG Cur Sp; TrG NSub Wr; Con Cur Sp
 - 2b. Ceiling is reached and maintained: TrG NSub Sp; TrG Rec Sp; TrG Def Wr; TrG Sub Wr; TrG Cur Wr; TrG Rec Wr
 3. Stabilization at a level below a ceiling: ConG NSub Sp
 4. Losses in target-like use
 - 4a. Loss in level: ConG NSub Wr
 - 4b. Loss in slope, stabilization: ConG Def Sp; ConG Sub Sp
 - 4c. Loss in slope, stabilization, gain in level: ConG Rec Sp
 - 4d. Losses in level and slope, stabilization: ConG Def Wr; ConG Cur Wr; ConG Rec Wr
 - 4e. Losses in level and slope, decreases in TLU: Con Sub Wr
-

In the data, there is one example the first category of changes from PreI phase to PostI phase, gains in target-like use, with gains in both level and slope: the treatment group's suppliance of *the* in subject position on the spoken task (gain in level: 26.4; gain in slope: .28; cf. Appendix J4).

The second category of changes from PreI phase to PostI phase, continued development of target-like use or maintenance of a very high target-like use, is sub-divided into two sub-categories: 2a) no change in level, with a continued positive slope; and 2b) ceiling is reached during the PreI phase and maintained during the PostI phase. In the data, there are four examples of sub-category 2a, no change in level, with a continued positive slope: the treatment group's suppliance of *the* for referential definites on the spoken task (no major change in level: 3.9; no major change in slope: .08; PreI slope: .33; PostI slope: .41; cf. Appendix J1); the treatment group's suppliance of *the* for current NPs on the spoken task (no major change in level: -2.5; no major change in slope: .11; PreI slope: .44; PostI slope: .55; cf. Appendix J8); the treatment group's suppliance of *the* for in non-subject positions on the written task (no major change in level: 0.2; no major change in slope: .04; PreI slope: .28; PostI slope: .32; cf. Appendix J7); and the control group's suppliance of *the* for current NPs on the spoken task (no major change in level: -7.2; no major change in slope: -.13; PreI slope: .35; PostI slope: .22; cf. Appendix J8).

In the data, there are six examples of sub-category 2b, ceiling is reached during the PreI phase and maintained during the PostI phase: the treatment group's suppliance of *the* in non-subject positions on the spoken task (estimated score for the first day of the PostI phase, based on the PreI regression line: 91.9; slope during the PostI phase: .40; gain in slope: .16; loss in level: 20.5; the loss in level followed an increase in slope would appear to indicate a minor fluctuation after instruction; cf. Appendix J6); the treatment group's suppliance of *the* for recent NPs on the spoken task (estimated score for the first day of the PostI phase, based on the PreI regression line: 94.7; slope during the PostI phase: .23; cf. Appendix J10); the treatment group's suppliance of *the* for referential definites on the written task (estimated score for the first day of the PostI phase, based on the PreI regression line: 92.8; slope during the PostI phase: .12; cf. Appendix J2); the

treatment group's suppliance of *the* in subject position on the written task (estimated score for the first day of the PostI phase, based on the PreI regression line: 132.4; slope during the PostI phase: -.03; cf. Appendix J5); the treatment group's suppliance of *the* for current NPs on the written task (estimated score for the first day of the PostI phase, based on the PreI regression line: 97.5; slope during the PostI phase: .08; cf. Appendix J9); and the treatment group's suppliance of *the* for recent NPs on the written task (estimated score for the first day of the PostI phase, based on the PreI regression line: 119.5; slope during the PostI phase: -.05; cf. Appendix J11).

The third category of changes from PreI phase to PostI phase, stabilization at a level below a ceiling, is characterized by a flat slope during the PostI phase, with the level of that flat slope not at a very high target-like level, that is, not at a ceiling. In the data, there is one example of this category: the control group's suppliance of *the* in non-subject positions on the spoken task (slope during the PreI phase: -.22; slope during the PostI phase: .04; day 85 prediction, based on the postRL: 26.9; cf. Appendix J6). Between the two phases, there was a gain of .26 in slope and 14.5 in level, and therefore this data set could also be characterized as a gain in slope and level. However, because the slope in the PostI phase was very close to .00, it was felt that placing this data within category 1a would obscure this apparent stabilization at a low level.

The fourth category of changes from PreI phase to PostI phase, losses in target-like use, is sub-divided into five sub-categories: 4a) loss in level; 4b) loss in slope, resulting in stabilization; 4c) loss in slope, resulting in stabilization, and with a small gain in level; and 4d) losses in level and slope, resulting in stabilization; and 4e) losses in level and slope, resulting in decreases in target-like use. In the data, there is one example of sub-category 4a, loss in level: the control group's suppliance of *the* in non-subject positions on the written task (loss in level: -13.7; no major change in slope: -.08; cf. Appendix J7).

In the data, there are two examples of sub-category 4b, loss in slope, resulting in stabilization: the control group's suppliance of *the* for referential definites on the spoken task (no change in level: 0.0; loss in slope: -.23; PostI slope: .08; cf. Appendix J1); and the control group's suppliance of *the* in subject position on the spoken task (no major change in level: 8.5; loss in

slope: $-.39$; PostI slope: $-.04$; cf. Appendix J4). In the data, there is one example of sub-category 4c, loss in slope, resulting in stabilization (with a small gain in level): the control group's suppliance of *the* for recent NPs on the spoken task (a small change in level: 18.4; major loss in slope: $-.32$; PostI slope: $-.10$; cf. Appendix J10).

In the data, there are three examples of sub-category 4d, losses in level and slope, resulting in stabilization: the control group's suppliance of *the* in all definite contexts on the written task (loss in level: -25.0 ; loss in slope: $-.54$; PostI slope: $.01$; cf. Appendix J2); the control group's suppliance of *the* for current NPs on the written task (loss in level: -45.0 ; loss in slope: $-.85$; PostI slope: $-.01$; cf. Appendix J9); and the control group's suppliance of *the* for recent NPs on the written task (loss in level: -37.6 ; loss in slope: -1.51 ; PostI slope: $.03$; cf. Appendix J11). In the data, there is one example of sub-category 4e, losses in level and slope, resulting in decreases in target-like use: the control group's suppliance of *the* in subject position on the written task (loss in level: -11.5 ; loss in slope: -1.24 ; PostI slope: $-.29$; cf. Appendix J5);

This summary of the regression lines allows an assessment of the effects of instruction, with the PostI scores of the treatment group compared to two sets of baseline scores: (a) their PreI baseline values, Hypothesis 7a; and (b) the scores of the control group, Hypothesis 7b.

5.2.2.1 Hypothesis 7a: Treatment Group, Gains, Regression Lines, All Dependent Variables, All Tasks

The predictions for effects of instruction within the treatment group were developed in section 2.5 and are repeated below.

Hypothesis 7a, regarding changes in developmental trend in the instructed group:
With respect to the five dependent variables on the written, spoken, and cloze tasks, the instructed group will exhibit continued or improved developmental trends from the PreI phase to the PostI phase.

As noted above, both groups performed at a ceiling level on the cloze task during both phases, and therefore those results are not considered further in this section. A comparison within the

treatment group reveals that the four instructed learners exhibited consistent improvement from the PreI phase to the PostI phase. For one of the dependent variables, the treated learners exhibited gains both slope and level. This dependent variable, *the* suppliance in subject position on the spoken task, is the aspect of *the* suppliance that was predicted to improve most clearly if this interlanguage knowledge was automatized. For the other nine variables, the treated learners either maintained the PreI developmental trend or maintained a very high target-like use, perhaps a ceiling, that had been attained during the PreI phase. Therefore, acting as their own controls, the treated learners demonstrated continued or improved development after instruction, supporting Hypothesis 7a and providing evidence for the beneficial effects of instruction.

5.2.2.2 Hypothesis 7b: Group Differences, Regression Lines, All Dependent Variables, All Tasks

The predictions for effects of instruction in both the treatment and control groups were developed in section 2.5 and are repeated below.

Hypothesis 7b, regarding group differences in changes in developmental trend:
With respect to the five dependent variables on the written, spoken, and cloze tasks, the instructed group will exhibit a greater number of positive changes in developmental trend from the PreI phase to the PostI phase than will the uninstructed group.

As noted above, both groups performed at a ceiling level on the cloze task during both phases, and therefore those results are not considered further in this section. An examination of the changes in regression lines for the control group permits a comparison between the two groups. For eight of the ten dependent variables, the control group exhibited some type of decrease in *the* suppliance during the PostI phase: loss of level or loss of slope (i.e. the cessation of development), or both. For only two dependent variables, *the* suppliance for current NPs on the spoken task (Appendix J8) and *the* suppliance in non-subject positions on the spoken task (Appendix J6) did the control group not exhibit a decrease: these PreI to PostI comparisons revealed (a) continued development with a slightly reduced level and slope, or (b) stabilization, respectively. Overall, these results indicate that the control group was improving during the PreI phase, but that they did not continue to improve during the PostI phase, either stabilizing or decreasing in their suppliance of *the*.

A comparison of the treatment group and the control group reveals considerable differences between the two groups and strongly supports Hypothesis 7b, suggesting beneficial effects of instruction. Whereas the treatment group demonstrated continued development or gains during the Post I phase, the control group generally exhibited losses. This sharp contrast between the two groups, differences much more extreme than those noted in section 5.1.3, may have resulted from instruction rather than simply from continuing developmental trends. Stabilization, which may result in fossilization, might have occurred for the instructed learners if they had not received instruction.

This between group comparison of these regression lines has a very important implication for interpreting the results reported in section 5.2.1, where group mean phase scores were used in the analyses. The regression lines revealed that the control learners generally improved during the PreI phase (cf. section 5.1), but that during the PostI phase they did not improve above the level that they had achieved by the end of the PreI phase. Thus, the regression lines indicated no development during the PostI phase. In contrast, the use of a group mean phase score suggested a difference between phases because the mean PreI phase score, as an average, was a value that was lower than the level attained by the end of the PreI phase. Therefore, because of the PreI developmental trend, the differences between the groups was obscured when a group mean phase score was used.

Although the use of the regression lines revealed differences between the groups, the interpretation of these between-group results regarding the positive effects of instruction must remain tentative. As noted in sections 4.1.1 and 5.1.3, the two groups were self-selected and non-equivalent in many respects. Stronger evidence for the beneficial long term effects of instruction, preventing fossilization, would require, as discussed in sections 2.3 and 3.3.1, an equivalent control group (cf. Mellow, Reeder & Forster, in press). However, the combination of the between-group differences and the within-group comparisons provides compelling evidence for the beneficial long- (or mid-) term effects of instruction, including effects on an unmonitored spoken task.

5.2.3 *Variation According to Linguistic Context and Task, PreI - Post I Changes, Group Mean Gain Scores and Regression Lines*

The analyses in the two previous sections have used group mean phase scores and regression lines to consider between and within group differences in development on each dependent variable. In particular, the information provided by the regression lines has provided evidence for beneficial effects of instruction. In this section, the dependent variables and tasks are not considered individually. Instead, the data are analysed in terms of the interaction of task (spoken and written) and linguistic context (Sub - NSub, and Rec - Cur), using both the group mean phase scores (with inferential statistical tests of these means) and group regression lines for the analyses. In section 5.1.2, an analysis of the data from the PreI phase indicated differences according to task and linguistic context, differences that were consistent with the principles outlined in section 2.4. In this section, developmental patterns in those differences are examined. These analyses are particularly powerful and revealing because they can provide an explanation of complex patterns of interlanguage use and development, and because they permit an examination of the effects of instruction with respect to specific areas of L2 use that are hypothesized to most clearly benefit from instruction, including a reduction in the efficient elimination of redundant items because of increases in automatization.

5.2.3.1 *Hypothesis 8a: Syntactic Position, Spoken Task, Treatment Group, Changes*

The predictions for effects of instruction on the efficient deletion of redundant elements on the spoken task were developed in section 2.5 and are repeated below.

Hypothesis 8a, regarding changes in the efficient deletion of redundant elements on the spoken task for the instructed group:

With respect to *the* suppliance in subject and non-subject positions on the spoken task, the instructed group will exhibit a reduction in the NSub > Sub asymmetry in the PostI phase.

The mean phase scores for the treatment group's suppliance of *the* in subject and non-subject positions on the spoken task are provided in Table 5.15.

Table 5.15
Mean Phase Scores for the Treatment Group for the Suppliance in Subject and Non-Subject Positions on the Spoken Task

	Sub	NSub	Sub - NSub
postI mean	74.3	84.8	-10.5
preI mean	30.0	77.5	-47.5
gain	44.3	7.3	37

The analysis of the PreI data, discussed in section 5.1.2.1 and indicated in Table 5.15, revealed a NSub > S asymmetry in speech. The mean phase scores in Table 5.15 indicate that during the PostI phase the NSub > Sub asymmetry in speech was reduced considerably, from a difference of 47.5% to a difference of 10.5%. These results are consistent with the hypothesized reduction in the efficient elimination of redundant elements because of greater automatization. The regression lines (Appendix J4 and J6; and in Table 5.13) indicate that in the PostI phase the rate of development increased for *the* suppliance in both Sub and NSub positions (increases in slope of .28 and .16, respectively), and that the loss of the asymmetry was due to a jump in level (26.4%) for *the* provision in subject position immediately following instruction, compared to a loss of level for non-subject positions (-20.5%), and partially due to the faster rate of development of *the* suppliance in subject positions (.10 greater for Sub position than NSub position during the PostI phase). Therefore, Hypothesis 8a is confirmed by these data, suggesting that knowledge of *the* suppliance was automatized as a result of instruction.

5.2.3.2 Hypothesis 8b: Syntactic Position, Spoken Task, Group Differences, Changes

The predictions for effects of instruction on group differences in changes in the efficient deletion of redundant elements on the spoken task were developed in section 2.5 and are repeated below.

Hypothesis 8b, regarding group differences in changes in the efficient deletion of redundant elements on the spoken task:

With respect to *the* suppliance in subject and non-subject positions on the spoken task, the instructed group will exhibit a greater reduction in the NSub > Sub asymmetry in the Post I phase than will the uninstructed group.

In contrast to the scores of the treatment group described above, the mean phase scores for the control group's suppliance of *the* in subject and non-subject positions on the spoken task are provided in Table 5.16.

Table 5.16
Mean Phase Scores for the Control Group for the Suppliance in Subject and Non-Subject Positions on the Spoken Task

	Sub	NSub	Sub - NSub
postI mean	41.3	48.3	-7
preI mean	5	44.3	-39.3
gain	36.3	4	32.3

The group mean phase scores indicate that, like the treatment group, the NSub > Sub asymmetry in speech was reduced considerably, from a difference of 39.3% to a difference of 7%. These results are consistent with the hypothesized reduction in the efficient elimination of redundant elements because of greater automatization, which might be attributed to the extensive exposure to English of the control learners. The reduction of the asymmetry for the treatment group (a phase difference of 37 percentage points) is slightly larger than the reduction of the asymmetry for the control group (a phase difference of 32.3 percentage points). However, both groups have comparable PostI phase NSub > Sub asymmetries (Tr: 10.5 percentage points; Con: 7 percentage points).

In order to assess the significance of the interaction between the gains of the treatment and control groups on the spoken task for *the* suppliance in subject and non-subject positions, a three factor doubly repeated measures ANOVA was performed on the PreI and PostI phase pooled scores of each individual (the group means and standard deviations were reported in Table 5.11 above). The null hypothesis of no Position x Time effect was rejected: a significant effect was found, $F = 23.58 (1, 6), p < .0028$. This result confirms that both groups exhibited a reduction in the NSub > Sub asymmetry on the spoken task. However, the null hypothesis of no Position x Time x Group effect was not rejected: a significant effect was not found, $F = 0.09 (1, 6), p = .78$ (n.s.). This result indicates that the slightly larger reduction of the efficiency asymmetry for the treatment group was not statistically significant: there were no group differences in this

development, as assessed by group mean phase scores. These results are similar to the results found in section 5.2.1 above, which indicated no group differences in the mean phase scores, and suggest that Hypothesis 8b is not confirmed.

However, an examination of the developmental graphs in Appendix J indicates that the group mean phase scores may not validly represent the actual developmental patterns of the control group. The regression lines (Appendix J4 and J6), as reported in Table 5.13, indicate that, whereas the treatment group showed an improved rate of development during the PostI phase, the control group exhibited stabilization of *the* suppliance in subject position (a decrease in slope of $-.39$, to a PostI slope of $-.04$) and stabilization for *the* suppliance in non-subject positions (an increase in slope of $.26$, to a flat PostI slope of $.04$). In fact, as indicated in section 5.1.2.1, for the control group the NSub > Sub asymmetry disappeared during the PreI phase, not during the PostI phase. In addition, an examination of the graphs (Appendix J4 and J6) indicates that the reduction in the NSub > Sub (priming) asymmetry was not the result of automatization (i.e., continuing or maintained improvement in *the* suppliance in subject position), but that the increase in the mean for *the* suppliance in subject position was due to unusual variability in the data. Specifically, there were two short time periods, one during days 78 and 85 and the other on day 173, when three of the control learners supplied *the* in subject position substantially above the levels that were exhibited at all other times during the study. The cause of this variability is not clear.²² While one hypothesis may be that this increased variability of *the* suppliance in subject position is evidence of a restructuring of the control learners' interlanguage systems, perhaps the most plausible explanation is that unsystematic variability is a fundamental property of L2 use (e.g., Ellis, 1985, 1994), and that this variability was less evident in the data from the instructed learners due to their higher proficiency levels and perhaps due to the influence of instruction, an issue that is further explored in section 5.3. Therefore, except for learner Con3, as discussed in

²² On days 78 and 85, two different fable versions were retold by the two learners with unusually high *the* suppliance. On day 173, the same fable version was retold by the two learners with unusually high *the* suppliance. Therefore, differences in fable version may have contributed to this variation.

section 5.3, the reduction in the NSub > Sub asymmetry in speech does not appear to be the result of greater automatization of knowledge of article use. In sum, Hypothesis 8b is partially supported by the regression lines, suggesting that knowledge of *the* suppliance was automatized for the instructed group and that knowledge of *the* suppliance may have become more automatized for one of the control learners, but that PreI to PostI changes for the control group may be due to unexplained variability in the data.

5.2.3.3 Hypothesis 9: Syntactic Position, Written Task, Both Groups, Changes

The predictions for the effects of instruction on changes in the priming effect of redundant elements on the written task for both groups were developed in section 2.5 and are repeated below.

Hypothesis 9, regarding changes in the priming effect of redundant elements on the written task for both groups:

With respect to *the* suppliance in subject and non-subject positions on the written task, both groups will exhibit a continued Sub > NSub asymmetry in the Post I phase. However, if one or both groups achieve a ceiling (near 100% suppliance), then the facilitating effects of priming may be obscured, with suppliance being equal in the two positions.

The treatment group, as discussed in section 5.1.2.2, progressed very rapidly during the PreI phase in their suppliance of *the* in subject position on the written task. As a result, the mean phase scores obscured a Sub > NSub asymmetry on the written task that developed before the end of the PreI phase. Therefore, the PostI mean scores of *the* suppliance in subject and non-subject positions on the written task are compared to the day 80 scores estimated by the regression lines, as in Table 5.17.²³

²³ When a three factor doubly repeated measures ANOVA was performed on the PreI and PostI phase pooled scores of each individual (the group means and standard deviations were reported in Table 5.11 above), neither the null hypothesis of no Position x Time effect ($F = 4.27 (1, 6), p = .08$ [n.s.]) nor the null hypothesis of no Position x Time x Group effect ($F = 1.28 (1, 6), p = .30$ [n.s.]) was rejected. Although these results suggest that there was not a significant change in the priming asymmetry between the phases and that the groups were similar in their changes (or lack of changes) in asymmetrical suppliance patterns, these results were not considered further because it was felt that the group mean phase scores did not validly represent the rapidly changing group performance during each phase (cf. section 5.1.2.2).

Table 5.17
Estimated Day 80 Scores (from the PreI Regression Lines) and Mean PostI Phase Scores for the Treatment Group for the Suppliance in Subject and Non-Subject Positions on the Written Task

	Sub	NSub	Sub - NSub
postI mean	93.3	90.8	2.5
day 80	121	81.4	18.6 ^a
gain	-6.7 ^a	9.4	-16.1

^a These scores were calculated after the predicted value had been adjusted to 100%.

These phase scores indicate that the Sub > NSub asymmetry on the written task was reduced, from a difference of 18.6% to a difference of 2.5%. These results are consistent with the ceiling effects hypothesized in section 2.4 (Principle 6). The regression lines (Appendix J5 and J7; and in Table 5.13) indicate that during the PostI phase *the* suppliance in subject position maintained the ceiling level achieved during the PreI phase, and that the loss of the asymmetry was largely due to continued development in *the* provision in non-subject positions during the PostI phase. In sum, the disappearance of the Sub > NSub asymmetry is consistent with the predicted ceiling effects, with the development toward the ceiling potentially due to automatization as well as due to metalinguistic knowledge that developed as a result of the instruction.

The control group, like the treatment group, progressed very rapidly during the PreI phase in its suppliance of *the* in subject position on the written task (see section 5.1.2.2). As a result, the mean phase scores partially obscured a Sub > NSub asymmetry on the written task that developed before the end of the PreI phase. Therefore, the PostI mean scores of *the* suppliance in subject and non-subject positions on the written task are compared to the day 80 scores projected by the regression lines, as in Table 5.18. These phase scores for the control group indicate that the Sub > NSub asymmetry on the written task remained large, decreasing from a difference of 56.6% to a difference of 31.8%. These results, with no evidence of ceiling effects, are consistent with the Priming Hypothesis of Mellow and Cumming (1994) and Hypothesis 9. In addition, these phase scores suggest that the control group's suppliance of *the* did not become more target-like during the PostI phase (with a loss of 16.7% for Sub and a small gain of 8.1% for NSub) (a confirmation

Table 5.18
Estimated Day 80 Scores (from the PreI Regression Lines) and Mean PostI Phase Scores for the Control Group for the Suppliance in Subject and Non-Subject Positions on the Written Task

	Sub	NSub	Sub - NSub
postI mean	83.3	51.5	31.8
day 80	105	43.4	56.6 ^a
gain	-16.7 ^a	8.1	-24.8

^a These scores were calculated after the predicted value had been adjusted to 100%.

of Hypothesis 7b). The regression lines for the PostI phase (Appendix J5 and J7; and summarized in Table 5.13) indicate that, whereas the treatment group stayed at a ceiling level or maintained development until reaching a ceiling level, the control group appeared to remain at a ceiling for *the* suppliance in subject position (with a slight decrease by the end of the PostI phase), but stabilized in *the* suppliance in non-subject positions, exhibiting a reduction in both level (-13.7%) and rate (-.08) of development. Overall, this analysis of the interaction of instruction (i.e., group), task, and syntactic position suggests that the instruction was effective and that the cognitive processing model outlined in Chapter 2 is able to effectively account for these systematic changes in interlanguage.

5.2.3.4 Hypothesis 10: Referential Distance, Spoken Task, Both Groups, Changes

As noted in section 2.5 and confirmed in section 5.1, a unique set of interactions is expected for instruction, task and referential distance. The predictions for the effects of instruction on the changes in the facilitating effect of polyfunctional elements on the spoken task for both groups were developed in section 2.5 and are repeated below.

Hypothesis 10, regarding changes in the facilitating effect of polyfunctional elements on the spoken task for both groups:

With respect to *the* suppliance for current and recent referents on the spoken task, both groups will exhibit a continued Rec > Cur asymmetry in the Post I phase. However, if one or both groups achieve a ceiling (near 100% suppliance), then the facilitating effects of polyfunctionality may be obscured, with suppliance being equal for both types of referents.

The mean phase scores for the treatment group's suppliance of *the* for recent and current noun phrases on the spoken task are provided in Table 5.19.

Table 5.19
Mean Phase Scores for the Treatment Group for the Suppliance for Recent and Current Noun Phrases on the Spoken Task

	Rec	Cur	Rec - Cur
postI mean	79.3	80.3	-1
preI mean	56.5	45.5	11
gain	22.8	34.8	12

The mean phase scores indicate that the small Rec > Cur asymmetry in speech was eliminated, from a difference of 11% to a difference of -1%. These results, the loss of an effect for referential distance, are consistent with the ceiling effects hypothesized in section 2.4 (Principle 6). As noted in section 5.2.1, the mean phase scores also indicate large gains between the two phases. The regression lines (Appendix J8 and J10; and in Table 5.13) indicate that the loss of the asymmetry was due to a decrease in the rate of development for *the* suppliance for recent NPs (a decrease in slope of .38, to a PostI slope of .23) and a slight increase for *the* suppliance for current NPs (an increase in slope of .11, to a PostI slope of .55). The apparent interpretation of these changes is that *the* suppliance for both current and recent NPs reached or maintained a ceiling during the PostI phase. The regression lines generally indicate continued development during the PostI phase, potentially suggesting that instruction may have prevented stabilization from occurring at a level below the ceiling.

The mean phase scores for the control group's suppliance of *the* for recent and current noun phrases on the spoken task are provided in Table 5.20.

Table 5.20
Mean Phase Scores for the Control Group for the Suppliance for Recent and Current Noun Phrases on the Spoken Task

	Rec	Cur	Rec - Cur
postI mean	48.5	38.8	9.8
preI mean	24	17.3	6.8
gain	24.5	21.5	3

The mean phase scores indicate that the small Rec > Cur asymmetry in speech increased slightly, from a difference of 6.8% to a difference of 9.8%. These results, the continuation of an effect for referential distance at levels below the ceiling, are consistent with Hypothesis 10. In addition, the gain in the mean phase scores suggest development during the PostI phase, although as discussed in section 5.2, the regression lines indicate that the control group did not improve during the PostI phase.

In order to assess the significance of the interaction between the gains of the treatment and control groups on the spoken task for *the* suppliance for current and recent NPs, a three factor doubly repeated measures ANOVA was performed on the PreI and PostI phase pooled scores of each individual (the group means and standard deviations were reported in Table 5.11 above). The null hypothesis of no Distance x Time effect was not rejected: a significant effect was not found, $F = 0.29 (1, 6), p = .61$ (n.s.). The null hypothesis of no Distance x Time x Group effect was also not rejected: a significant effect was not found, $F = 0.85 (1, 6), p = .39$ (n.s.). These results suggest (a) that there was not a significant change in *the* suppliance patterns with respect to referential distance between the phases, a result expected because both the principles affecting this pattern are not affected by automatization, and (b) that the groups were similar in their changes (or lack of changes) in suppliance patterns.

As noted in previous sections, the use of group mean phase scores appears to have obscured certain aspects of the nature of the developmental data. Just as the data for the control group's suppliance of *the* in subject position on the spoken task did not reveal a systematic pattern of development, but instead exhibited unusual variability (section 5.2.3.2), an examination of the regression lines for the control group's suppliance of *the* for current and recent referents on the spoken task also reveals considerable variation. Thus, it was already noted in section 5.1.2.3 that the Rec > Cur asymmetry disappeared for the control group by the end of the PreI phase, changing to a Cur > Rec asymmetry of 4.4% by day 80. In addition, the regression lines for the PostI phase (Appendix J8 and J10; and in Table 5.13) suggest that, at the beginning of the PostI phase, there was a large Rec > Cur asymmetry, an estimated difference of 20.5% (at day 85: *the*

suppliance for Rec: 39.8%, *the* suppliance for Cur: 19.3%). This difference is larger than the asymmetry indicated by the PostI mean phase scores and opposite of that predicted by the PreI regression lines for day 85. In addition, by the end of the PostI phase, the PostI regression lines indicate that the asymmetry was reversed, with *the* suppliance for current NPs 9.9% higher than for recent NPs (at day 180: *the* suppliance for Rec: 30.3%, *the* suppliance for Cur: 40.2%). This change was largely due to a major shift in developmental trend for *the* suppliance for recent NPs, a switch from steady improvement to a decrease in *the* suppliance (PreI slope: .22, PostI slope: -.10, change: -.32). The rate of improvement of *the* suppliance for current NPs decreased during the PostI phase, but continued to exhibit a positive trend (PreI slope: .35, PostI slope: .22, change: -.13). Therefore, the regression lines for the control group for both phases indicate continued fluctuations in *the* suppliance for these linguistic contexts, and do not indicate an overall trend toward greater *the* suppliance during the PostI phase. Thus, while the instructed group continued developing toward or maintaining target-like use, for the control group during the PostI phase, development appeared to stop, with continued fluctuations of suppliance in these contexts. Although the fluctuations in the referential distance asymmetry may suggest that some type of restructuring was occurring, the most plausible explanation, as noted in section 5.2.3.2 and further explored in section 5.3, is that the control exhibited continued variation in L2 use while the instructed learners exhibited increasingly systematic and consistent L2 use. Although the groups are somewhat non-equivalent, these differences are nonetheless suggestive of the beneficial effects of instruction.

Overall, the mean phase scores for the control group are consistent with polyfunctionality effects, but the regression lines indicate that these mean scores obscured fluctuations in *the* suppliance during the PostI phase. In addition, as discussed in section 5.2, the group mean phase scores suggested similar treatment and control group gains, but the regression lines indicated that only the treatment group continued improving during the PostI phase. Thus, Hypothesis 10 is partially confirmed.

5.2.3.5 Hypothesis 11: Referential Distance, Written Task, Both Groups, Changes

The predictions for the effects of instruction on changes in the facilitating effect of polyfunctional elements on the written task for both groups were developed in section 2.5 and are repeated below.

Hypothesis 11, regarding changes in the facilitating effect of polyfunctional elements on the written task for both groups:

With respect to *the* suppliance for current and recent referents on the written task, both groups will exhibit a continued Rec > Cur asymmetry in the Post I phase. However, if one or both groups achieve a ceiling (near 100% suppliance), then the facilitating effects of polyfunctionality may be obscured, with suppliance being equal for both types of referents. In addition, these polyfunctionality effects may be obscured on the written task by the use of metalinguistic knowledge or other conscious strategies.

The mean phase scores for the treatment group's suppliance of *the* for recent and current noun phrases on the written task are provided in Table 5.21.

Table 5.21
Mean Phase Scores for the Treatment Group for the Suppliance for Recent and Current Noun Phrases on the Written Task

	Rec	Cur	Rec - Cur
postI mean	94.5	92.3	2.3
preI mean	68.5	67.8	0.8
gain	26	24.5	1.5

The mean phase scores indicate that there was no effect for referential distance in writing during either phase, with differences of 0.8% and 2.3%, respectively. These results would appear to be consistent with the hypothesis that interlanguage systems will be less evident on the written task, potentially due to the obscuring influence of conscious knowledge (Principle 5). As noted in section 5.2.1, the mean phase scores also indicated large gains between the two phases. The regression lines (Appendix J9 and J11; and in Table 5.13) also generally indicated the absence of an asymmetry, with *the* suppliance for both types of referents developing steadily during the PreI phase, and then reaching or maintaining a ceiling during the PostI phase. However, *the* suppliance for recent referents began lower and developed faster than for current referents, reaching a ceiling first, and this may have resulted in a small Rec > Cur asymmetry near the end of the PreI phase.

Because there was not an large asymmetry in the PreI phase, the effects of instruction appear to have been only those discussed in section 5.2.2: stabilization at a ceiling throughout the PostI phase for both contexts.

The mean phase scores for the control group's suppliance of *the* for recent and current noun phrases on the written task are provided in Table 5.22.

Table 5.22
Mean Phase Scores for the Control Group for the Suppliance for Recent and Current Noun Phrases on the Written Task

	Rec	Cur	Rec - Cur
postI mean	59.8	71.8	-12
preI mean	63.5	59.5	4
gain	-3.8	12.3	16

The mean phase scores indicate that the very small Rec > Cur asymmetry in the written task was lost and reversed during the PostI phase, changing from 4% to -12%. These results are not consistent with Hypothesis 11, but are consistent with the manner in which the control group developed on the spoken task, with alternating asymmetries throughout the 6 month period for these linguistic contexts (cf. section 5.2.3.4). The regression lines (Appendix J8 and J10; and in Table 5.13) indicate that there was considerable development in *the* suppliance for both types of referents during the PreI phase, suggesting that a ceiling was reached by the end of that phase (with a slight Rec > Cur asymmetry possible near the end of the PreI phase). However, both variables exhibited a considerable loss in level and flattening of slope during the PostI phase (for *the* suppliance with recent NPs, loss in level: -37.6; loss in slope: -1.51; for *the* suppliance with current NPs, loss in level: -45.0; loss in slope: -.85). In contrast to the mean phase scores which indicated a small Cur > Rec asymmetry during the PostI phase, the regression lines indicate a small Rec > Cur asymmetry during the PostI phase. This conflicting information between the mean phase scores and the regression lines appears to indicate both that referential distance did not have a strong effect in these data and that the control learners may have been particularly variable in their L2 use.

In order to assess the significance of the interaction between the gains of the treatment and control groups on the written task for *the* suppliance for current and recent NPs, a three factor doubly repeated measures ANOVA was performed on the PreI and PostI phase pooled scores of each individual (the group means and standard deviations were reported in Table 5.11 above). The null hypothesis of no Distance x Time effect was not rejected: a significant effect was not found, $F = 0.90 (1, 6), p = .38$ (n.s.). The null hypothesis of no Distance x Time x Group effect was also not rejected: a significant effect was not found, $F = 1.31 (1, 6), p = .30$ (n.s.). These results suggest (a) that there was not a significant change in *the* suppliance patterns with respect to referential distance between the phases, a result expected because the principles affecting this pattern are not affected by automatization, and (b) that the groups were similar in their changes (or lack of changes) in suppliance patterns. However, as noted above, the use of group mean phase scores appears to have obscured trends in the developmental data.

These results suggest two interesting findings for referential distance effects. The variability of L2 use by the control group is the first interesting finding with respect to referential distance. On both the spoken and written tasks and throughout the duration of the study, *the* suppliance by the control group exhibited minor fluctuations in the asymmetrical relations with respect to recent and current referents. The issue of variability will be explored further in section 5.3.

Second, as noted in section 5.2.2.2, the regression lines indicate that, whereas the treatment group remained stable at a ceiling level for both contexts on the written task during the PostI phase, the control group exhibited considerable loss of level and rate of development during the Post I phase, apparently stabilizing well below the level of the treatment group. Because the control group had actually developed more rapidly than the treatment group during the PreI phase, the two groups appeared to be relatively equal by the end of that phase, and therefore the stabilization at a lower level contrasts noticeably with the treatment group. Although the groups were non-equivalent, it appears that the instructed group continued its target-like use, but the

uninstructed group either stabilized in its development or underwent some type of restructuring of their interlanguage systems.

Overall, the results of the analyses in this section have indicated that the hypotheses presented in Chapter 2 are able to account for interesting asymmetries and developments in these data. Particularly interesting is the finding that the instructed group lost its $N_{Sub} > Sub$ (efficiency) asymmetry in speech, exhibiting considerable gains in level and rate of development for *the* suppliance in subject position. In contrast, the control group exhibited a reduced rate of development for *the* suppliance in subject position during the PostI phase, stabilizing at a level far below target-like use. This important finding provides evidence which suggests that instruction can lead to one of the most desirable pedagogical effects, the automatization of interlanguage knowledge.

5.3 The Developmental Course of Instructed SLA: An Examination of Individual Development

The previous two sections have examined group behavior, that is, considered and compared the behavior of the two groups, control and treatment. This section will examine the individual behavior of the eight learners in the study. An examination of the developmental course will be achieved through a visual examination of graphs that represent the development of the eight learners.

5.3.1 Hypothesis 12: Developmental Course, Individual Patterns

The predictions for individual variation in developmental course were developed in section 2.5 and are repeated below.

Hypothesis 12, regarding the patterns of individual learners with respect to the developmental course of instructed SLA:

With respect to the five dependent variables on the written, spoken, and cloze tasks, each instructed and uninstructed learner will exhibit unique patterns of development, possibly including gradual improvement, immediate improvement, no improvement, and delayed attrition.

To address Hypothesis 12, this section considers examples of four major patterns in the developmental course of instructed SLA: 1) gradual improvement, 2) immediate improvement, 3) no improvement, and 4) delayed attrition. This discussion refers to the line graphs provided in Appendix B.

The first major pattern to be considered is gradual improvement in response to instruction. The data provides examples of two types of gradual improvement, without or with U-shaped development. The third treated learner (Tr3; Appendix B3), who chose to receive two sessions of instruction, provides an example of gradual development without a U-shaped pattern. This pattern is most clearly seen in the line graph of *the* suppliance for referential definites on the spoken task: after fluctuating considerably but showing little improvement during the PreI phase, learner Tr3 gradually improved during the PostI phase until reaching 100% at the end of the study. An examination of the scores in the linguistic sub-contexts reveals that prior to instruction and after one session of instruction, Tr3 exhibited a considerable NSub > Sub asymmetry, a pattern consistent with the efficient omission of redundant elements predicted by Principle 3. After the second session, there was an immediate improvement in *the* suppliance in subject position on the spoken task, a development consistent with the automatization predicted by Principle 4. This was then followed by fluctuations that cumulatively resulted in gradual overall improvement and eliminated the NSub > Sub asymmetry as the ceiling was reached, as predicted by Principle 6. Interestingly, an examination of current and recent NPs in the PostI phase reveals a slight but consistent Cur > Rec asymmetry on the spoken task (in opposition to the effects of polyfunctionality predicted by Principle 2), with *the* suppliance developing gradually in both contexts (Tr3's suppliance on the written task was 100%, with the symbols for Cur hidden by the symbols for Rec).

An example of U-shaped gradual development can be seen in the line graphs of learner Tr1 (Appendix B1), who chose to receive one session of instruction. The U-shaped curves are clear in the graphs for *the* suppliance for referential definites: After gradually improving during the PreI phase, immediately after instruction Tr1's *the* suppliance decreased on all tasks. Although this

decrease continues on the spoken task for the next session, Tr1 then gradually improves again, eventually attaining levels higher than those achieved during the PreI phase.

U-shaped development often indicates two stages of restructuring of linguistic knowledge, such as the change from a target-like exemplar based system (e.g., a memorization of the past tense form *went*) restructured to a rule-based system (e.g., production of the form *goed*) and then restructured again to include exceptions to the rule (e.g., production of *went*) (cf. McLaughlin, 1990). However, rather than restructuring, an interpretation of the U-shaped development of Tr1 appears to be consistent with automatization and a temporary uncertainty in the interlanguage system, an uncertainty which may or may not have been conscious. An examination of *the* suppliance according to syntactic position on the spoken task supports this interpretation. During the PreI phase there was a very clear and consistent NSub > Sub asymmetry on the spoken task, a pattern consistent with the efficient omission of redundant elements predicted by Principle 3: *the* suppliance in subject position averaged 4% and never rose above 9%, and *the* suppliance in non-subject positions averaged 86%, never fell below 71% and was as high as 100%. However, after instruction, *the* suppliance in non-subject position dropped considerably, with provision in non-subject position only 45% and 60% during the next two data collection sessions. The learner then improved considerably, increasing to 75% suppliance and achieving 100% in non-subject positions in the final three data collection sessions. Corresponding to the return to high levels of *the* suppliance in non-subject positions, Tr1 exhibited considerable improvement in subject position on the spoken task, achieving 67% and 63% suppliance in the final two data collection sessions. This gradual improvement in subject position is consistent with the automatization of redundant form-function mappings predicted by Principle 4.

This analysis suggests that the temporary loss in non-subject positions may have been due to temporary uncertainty (a change in use) rather than to a change in the interlanguage knowledge system. Such a temporary uncertainty can be interpreted in several ways. First, it could be claimed that the instruction temporarily confused and hindered the learner, but that the learner managed to overcome the distraction and return to PreI phase performance. Such an interpretation

would suggest that the form-focused instruction had a negative effect (e.g., Pienemann, 1989, pp. 72-76; cf. Ellis, 1994, pp. 652-654). However, an alternative interpretation that is better supported by the data is that the temporary uncertainty represents a period of learning caused by the instruction. During this period, the learner may have been selectively attending to article use in input and output as a result of the instruction (e.g., Ellis, 1994, pp. 657-658; Schmidt & Frota, 1986). This attention may have resulted in a reduction in *the* suppliance as the learner attended to and reconsidered article use. Although there had been no trend indicating development during the PreI phase for *the* suppliance in subject position on the spoken task, after this period of uncertainty the learner exhibited considerable improvement, with *the* suppliance becoming closer to *the* suppliance on the tasks that encouraged greater attention to form. This interpretation of a positive effect of instruction is supported by the group effects reported in section 5.2. In sum, the data on these individual line graphs illustrate both the group findings and the interlanguage patterns of gradual improvement, without and with U-shaped development.

The second major pattern that is exhibited in the data is immediate improvement. Two different types of immediate development can be identified: an immediate reduction in variability (the development of consistency) and an immediate increase in level of *the* suppliance. Learners Tr2 and Tr4 (Appendix B2 and B4), who chose to receive two and three sessions of instruction, respectively, provide examples of development in terms of instruction appearing to result in the development of consistency in interlanguage system use (cf. sections 5.2.3.2, 5.2.3.4, and 5.2.3.5). As is indicated by the development for all variables, learner Tr2 exhibited considerable variability during the PreI phase, sometimes performing near 100% and other times supplying *the* in only 20 - 40% of the contexts. After instruction, Tr2 immediately began to supply *the* in a target-like fashion and near the end of the study consistently supplied *the* in 100% of the contexts. Similar to learner Tr2, Tr4 also exhibited considerable variability during the PreI phase, although in the last two sessions prior to instruction Tr4 may have exhibited a somewhat more stable behavior. After instruction, Tr4 exhibited some variability (more than Tr2), but that variability was considerably less than that seen in the PreI phase. Both of these learners, Tr2 and Tr4, had

demonstrated near target-like behavior during the PreI phase and therefore their improvement seems to be more a matter of developing consistency than achieving an increase in level of *the* suppliance.

A second type of immediate development, an immediate increase in level during the PostI phase, is exhibited by learner Tr3 (Appendix B3). This pattern is clearly seen in Tr3's *the* suppliance for referential definites on the written task. Prior to instruction, Tr3 exhibited considerable variability with the scores averaging 59%. However, after the first session of instruction, Tr3 immediately supplied *the* in 100% of the written contexts and maintained a very high level for the following three fable sessions. A second example of an increase in level is exhibited by one of the control learners, Con3 (Appendix B7). During the PreI phase, Con3's scores were characterized by considerable variability and generally did not exceed 60%. However, during the PostI phase (but not after instruction), Con3's scores were almost all above 60%. Both examples of an immediate increase in level also exhibit increased consistency in scores, although this is difficult to assess for Con3 because only two sessions of data were collected during the PostI phase.

The third major pattern exhibited in these data is that of no improvement during the Post I phase. All of the members of the treatment group were described above because they exhibited some type of improvement. Only one member of the control group, Con3, exhibited PostI development. Thus, the other three control learners, Con1, Con2 and Con4 all exhibited no improvement during the PostI phase (Appendix B5, B6, and B8). All three control learners exhibited considerably variability through the study, sometimes achieving very high scores and other times supplying *the* in 0% of the contexts.

Delayed attrition is the fourth and final major pattern that is illustrated with these data. Delayed attrition can be seen in the scores of three of the treated learners, Tr1, Tr3 and Tr4 (Appendix B1, B3, and B4). For all of the learners, loss appears most clearly on the final day of data collection and, interestingly, is only on the written task (loss on the cloze task is not considered because it involves the omission of only one article by each learner). Unfortunately,

these results are confounded by task version because all four learners retold the same fable version during their final session, suggesting that the specific fable version may have been the cause of the slightly lower scores, although it is not clear why the changes occurred only on the written task. An examination of the linguistic contexts on the written data reveals that the loss was primarily in subject positions and for recent NPs, which is surprising because it was predicted that those linguistic contexts facilitate *the* suppliance, due to priming or polyfunctionality. Because there is no loss on the spoken task, this suggests that instruction may have had lasting effects on relatively unmonitored language use.

In sum, this examination of the line graphs of the individuals has illustrated a number of theoretically postulated patterns in the developmental course of SLA. Generalizations are not warranted from this data, due to small number of measurement sessions per individual, the small number of minimum contexts (three) required to report a session score, the relatively small scope of form-function mappings that were considered, and overall variation in the data. Nonetheless, the following observations can be made. First, interlanguage is characterized by considerable variation, especially in behavior that is not close to target-like. This is one reason why a large sample is required to get statistically significant results in an experimental study, and why, as noted above, it is difficult to generalize from a group to individuals. Certain aspects of this variation may be reduced by instruction, indicating an additional beneficial effect of instruction (cf. sections 5.2.3.2, 5.2.3.4, and 5.2.3.5). Second, in spite of this variation, a number of the principles proposed in section 2.4 were supported by these data. Third, it seems clear that individuals respond quite differently to instruction; improvement may be gradual or immediate, with or without U-shaped development. Fourth, because two out of the four instructed learners exhibited gradual improvement on the spoken task after instruction, this suggests that delayed post-tests are especially necessary to assess development of unmonitored language use. Delayed post-tests may also be important because of the possibility of U-shaped development: immediate post-tests may even suggest that instruction had negative effects. Overall, the data exhibited

interesting developmental patterns, varying according to task and linguistic context, and also indicated a number of important concerns regarding generalizability and research methodology.

Chapter 6

Limitations, Summary, and Implications

This chapter begins with a summary of the limitations of this study and then summarizes the findings in the previous chapter, indicating the relevance of those findings for SLA theory and L2 instruction. The chapter concludes with implications for pedagogy, theory and research.

6.1 Limitations

Throughout this study, attempts were made to achieve a high degree of pedagogical, linguistic, psycholinguistic, and internal validity. A fundamental controversy in L2 pedagogy, the value of form-focused instruction, was investigated. Instructional materials were designed to provide optimal sessions of instruction, with input processing and output practice activities informed by current theoretical and empirical findings. The topic of instruction was chosen because of pedagogical need: English article use is well-known as a difficult area for learners, especially Japanese learners. The analysis of article use was informed by a detailed analysis of the function of articles in English discourse and by a detailed and principled analysis of how articles might function in interlanguage discourse for these learners. In particular, an analysis of the efficient reduction of redundancy was utilized so that the potential effects of automatization could be assessed. The research design was carefully chosen to permit an investigation of development over time on a range of different tasks. Attempts were made to avoid the Hawthorne Effect, not drawing the learners' attention to articles, so that they might provide samples of language in which their attention was not focused on article use. Parallel tasks were designed to vary attention to form and to provide a relatively consistent task from session to session. In an effort to increase internal validity, postintervention language use was compared to both preintervention language use and the language use of a control group. In sum, the dissertation sought to address the underlying assumptions and design limitations of research into instructional outcomes, to outline linguistic and

cognitive processes that may account for L2 use and development, and then to determine the results of a study of these issues.

In spite of the care that was taken with these many aspects of design, the study was limited in a number of respects. The amount of instruction, one to three hours, and the distribution of instruction, confined to a short period and not recycled through a course or year, might not have been optimal for effecting maximal development. The type of instruction, individual tutoring, might not be generalizable to similar instruction provided for an entire class. The narrow focus of linguistic analysis, *the* suppliance for referential definites, missed many aspects of article use and many aspects of referent tracking in discourse. The methods by which the dependent variables were operationalized may not have been ideal, especially in terms of assessing referential distance by counting clauses rather than, for example, intonation units (Chafe, 1987). The analyses, although designed to be representative of redundancy and polyfunctionality mappings, may have had specific properties that limit their generalizability. The constructs of redundancy and polyfunctionality mappings, and of priming and efficient elimination, may be of limited utility for characterizing interlanguage systems. The sample size, four learners per group, and the number of data samples per phase, two to six, were too small to have strongly confirmed these findings and required that the results of the parametric statistical tests were interpreted with caution. The parallel versions of the fable retelling were not entirely parallel and were not completely counterbalanced, and therefore reliability was potentially affected by factors such as variation in the number of referents in a narrative. The limited number of elicited contexts for specific dependent variables resulted in scores reported when only three contexts of a variable were present, also limiting reliability. The tasks did not involve authentic, spontaneous language use, but instead were laboratory tasks that, at best, may have elicited language that was not often monitored for article use. And finally, the non-equivalence of the self-selected control and treatment groups made comparisons between groups suggestive rather than conclusive. As a result of this non-equivalence, the strongest claims for causality were achieved when the learners acted as their own controls. However, this comparison was problematic because of the developmental trend during

the preI baseline, and therefore it is not possible to know whether the instructed learners would also have stabilized during the postI phase.

Similar or comparable limitations could be cited for all instructional studies (cf. Beck & Eubank, 1991; Harley, 1989a, p. 335; Tomasello & Herron, 1991). I would argue, in fact, that such limitations are inescapable in the study of human behavior. Therefore, these limitations have two important implications. First, it is only through replications that greater confidence in findings can be achieved. Second, the results of specific studies must be interpreted with caution and in reference to theory. This cautious interpretation is provided in the next section.

6.2 Summary of Findings

The findings in this study are summarized with respect to the 12 hypotheses that were developed in section 2.5. The first five hypotheses were concerned with language use during the PreI phase. The results with respect to each hypothesis are summarized in Table 6.1. The table also indicates whether the principles (Ps) developed in section 2.4 were supported.

Table 6.1
Results of Hypotheses 1 Through 5

<u>Hypothesis</u>	<u>Results</u>
1. Task Variation, PreI Phase	Hypothesis supported
2. Syntactic Position, Spoken Task, Redundancy: Efficiency	Hypothesis supported, P3 (redundancy: efficiency) supported
3. Syntactic Position, Written Task, Redundancy: Priming	Hypothesis supported, P3 (redundancy: priming) supported, for predicted day 80 scores, by regression lines
4. Referential Distance, Spoken Task, Polyfunctionality	Hypothesis supported, P2 (polyfunctionality) supported
5. Referential Distance, Written Task, Polyfunctionality, Variation	Hypothesis supported, P2 (polyfunctionality) not supported, Task P5 (task variation) supported

As Table 6.1 indicates, all five hypotheses regarding task variation and linguistic context were supported, validating the set of linguistic and cognitive principles proposed in section 2.4.

Hypothesis 3 was supported by a detailed analysis of developmental trends during the PreI phase.

This analysis indicated the limitations of using group mean phase scores to assess language use and development. With respect to Hypothesis 5, polyfunctionality effects were not found on the spoken task, but the absence of these was argued to be consistent with the claim made in Principle P5, that, following Tarone (1988), systematic interlanguage patterns would be exhibited more clearly on the spoken task than on the written task. In addition, the language use of the two groups during the PreI phase was carefully compared, indicating a number of similarities and differences.

Hypotheses 6 through 11 were concerned with the effects of instruction. The results with respect to each hypothesis are summarized in Table 6.2. The table also indicates whether the principles (Ps) developed in section 2.4 were supported.

Table 6.2
Results of Hypotheses 6 Through 11

<u>Hypothesis</u>	<u>Results</u>
6. Group Differences, Group Mean Phase Scores, All Dependent Variables, All Tasks	Hypothesis not supported, Descriptive statistics supported trends for 9 of 11 comparisons Inferential statistical tests did not support the trends for any comparisons
7a. Treatment Group, Gains, Regressions Lines, All Dependent Variables, All Tasks	Hypothesis supported, beneficial effects of instruction suggested
7b. Group Differences, Regressions Lines, All Dependent Variables, All Tasks	Hypothesis supported, beneficial effects of instruction suggested, to the extent that the groups are equivalent
8a. Syntactic Position, Spoken Task, Treatment Group, Changes	Hypothesis supported, P3 (redundancy: efficiency) supported, P4 (automatization) supported, beneficial effects of instruction suggested
8b. Syntactic Position, Spoken Task, Group Differences, Changes	Hypothesis partially supported, P4 (automatization) partially supported, by unusual variability in ConG scores beneficial effects of instruction suggested by regression lines
9. Syntactic Position, Written Task, Both Groups, Changes	Hypothesis supported, by regression lines, P3 (redundancy: priming) supported, P6 (ceiling effect) supported, beneficial effects of instruction suggested by regression lines

Table 6.2 (continued)
Results of Hypotheses 6 Through 11

<u>Hypothesis</u>	<u>Results</u>
10. Referential Distance, Spoken Task, Both Groups, Changes	Hypothesis partially supported, P2 (polyfunctionality) partially supported, by mean phase scores P6 (ceiling effect) supported, beneficial effects of instruction suggested by regression lines
11. Referential Distance, Written Task, Both Groups, Changes	Hypothesis partially supported, P2 (polyfunctionality) not supported, P5 (task variation) supported beneficial effects of instruction suggested by regression lines

As Table 6.2 indicates, Hypothesis 6 was not supported; Hypotheses 8b, 10 and 11 were partially supported; and Hypotheses 7a, 7b, 8a, and 9 were supported. Hypothesis 6, a comparison of group gains, was confirmed by the group mean gain scores, but these differences were not found to be statistically significant. However, as indicated previously, it was determined that group mean phase scores provided only a crude measure of language use in a phase, and were inaccurate if there were substantial developmental trends or large differences in group trends. The value of instruction was therefore more accurately assessed by Hypotheses 7a and 7b, using the regression lines for within and between group comparisons, respectively. It was found that the treatment group exhibited continued or improved development during the PostI phase on both the spoken and written tasks, and that the control group exhibited stabilization or attrition during the same time period. Therefore, both within and between group comparisons indicated the beneficial effects of the instruction sessions.

The five hypotheses regarding the interaction of instruction, task, and linguistic context supported the beneficial effects of instruction and the set of linguistic and cognitive principles proposed in section 2.4. The beneficial effects of instruction were supported by (a) equivalence or increases in PostI regression lines for the treatment group, contrasted with losses in slope for the control group (a confirmation of the findings for Hypothesis 7b), and (b) the clear reduction of the NSub > Sub asymmetry on the spoken task for the treatment group, a finding that is effectively

explained as automatization of a redundancy mapping, resulting in the elimination of the efficient omission of *the* in subject position (Hypotheses 8a and 8b). The automatization of the efficiency asymmetry is perhaps the most important and convincing finding of the study, validating both the value of instruction and the linguistic and cognitive principles.

The validity of the linguistic and cognitive principles was further supported by the results regarding Hypotheses 9, 10, and 11. In particular, the principles regarding the priming of redundant elements on the written task (P3), task variation (P5), and ceiling effects (P6) were supported by the data. Although the facilitating effect of polyfunctionality was partially supported by the control group mean phase scores on the spoken task, these effects were not as clear as during the PreI phase, due to unexplained developmental trends.

Finally, Hypothesis 12 was concerned with the individual nature of trends in developmental course. An examination of the line graphs in Appendix B illustrated four patterns of development: 1) gradual improvement, with and without U-shaped development, 2) immediate improvement, as a reduction in variability and as an increase in level, 3) no improvement, and 4) delayed attrition. Only one control learner exhibited any type of improvement, whereas the four treated learners exhibited a variety of developmental patterns, thus confirming the group results found in section 5.2. The patterns that were illustrated also supported the interacting linguistic and cognitive principles, demonstrating principles of efficient omission of redundant elements (P3), automatization (P4), and ceiling effects (P6). The assessment of Hypothesis 12 also indicated a number of important concerns regarding generalizability and research methodology. In particular, the graphs revealed that interlanguage use was characterized by considerable variation and that individuals varied considerably in their responses to instruction. This variability suggests that it is difficult to generalize from group effects to individual language use, that an understanding of variability is essential for progress in SLA theory, and that these variations must be considered in research design, for example, with respect to the importance of delayed post-tests for assessing the effects of instruction.

6.3 Implications and Future Directions

A number of implications for pedagogy, SLA theory, and research methodology can be drawn from this study. First, because of the apparent beneficial effects of instruction on the written and spoken tasks, effects that generally continued for the eight to ten weeks of the study after instruction (cf. section 5.3), and because of the apparent automatization of interlanguage knowledge as revealed on relatively unmonitored tasks and in linguistic contexts that were predicted to reveal automatization (i.e., the loss of the efficient omission of redundant elements), it would be premature at this time to abandon form-focused instruction or to adopt a strictly emergent view of second language learning that does not acknowledge the possibility of the automatization of learned knowledge. Indeed, it is pedagogically meaningful that all of the instructed learners improved or continued in their development, whereas only one of the four uninstructed learners continued developing: Form-focused instruction may prevent fossilization. In addition, as argued by Breen (1987) with respect to a process syllabus that attempts to accommodate each learner's internal syllabus (cf. section 1.2), form-focused instruction may be particularly beneficial when learners deliberately choose to receive or utilize that instruction, as demonstrated by the self-selected groups in this study. Two important directions for future pedagogical studies are to identify those aspects of language that may be best facilitated through form-focused instruction, and to identify those properties of instruction (e.g. amount, timing, combinations of input processing and output practice) that best facilitate lasting changes in interlanguage knowledge.

Second, with respect to SLA theory, this study has indicated the complex interaction of the many factors that affect second language use. In section 1.4, three factors affecting variation were introduced: function, task, and time. These three factors were elaborated in chapter 2 to motivate the research hypotheses that guided this study. The interaction of these three factors in the principles that motivated the 12 research hypotheses is illustrated in Table 6.3.

Table 6.3
Principles Regarding the Interaction of Three Factors that Affect Variation in L2 Use

Principles	Function	Task	Time
Task Variation: Sp > Wr > Cl, Hypothesis 1 (cf. section 2.2)		√	
Principle 1: Topic-Comment Hypotheses 2, 3, 8a, 8	√		
Principle 2: Polyfunctionality Hypotheses 4, 10	√		
Principle 3: Redundancy Hypotheses 2, 3, 8a, 9	√	√	
Principle 4: Automatization Hypotheses 8a, 8b	√	√	√
Principle 5: Task Variation Hypotheses 5, 11		√	
Principle 6: Ceiling Effect Hypotheses 9, 10			√
Developmental Course Hypothesis 12 (cf. section 2.3)		√	√

As Table 6.3 indicates, these three factors interact in a complex manner. In spite of the methodological limitations noted above and throughout this dissertation, the consistency of the support for the research hypotheses can be interpreted as a preliminary validation of the proposed linguistic and cognitive principles of interlanguage use and development. A series of replication studies would strengthen the support for these findings. In particular, such replications could consider other redundancy and polyfunctionality mappings in interlanguage. Although such replications might focus on the expression of grammatical information such as number, tense, aspect, and old information, studies might also consider the communicative functions (e.g., requesting clarification, expressing social roles) of various discourse structures (e.g., repetition, topic shifts, intonation). In addition, the use of such studies to identify and validate additional interacting linguistic and cognitive factors will contribute to the development of SLA theory.

Third, given the qualified success of the methodology used in this study, and given the methodological limitations of this study, a number of which are identified in section 6.1, the methodological options and issues discussed in this dissertation suggest fruitful avenues for future research. In particular, the importance of longitudinal studies has been illustrated with these data, in terms of (a) having learners act as their own controls, (b) providing a large number of data points to validly assess development and permit the use of sophisticated statistical tests, (c) providing an interlanguage rather than target-language orientation, and (d) allowing an assessment of gradual and delayed effects of instruction. In addition, this study has indicated the importance of further studies addressing the relationship between group behavior and individual behavior, examining how generalizations might be made from group studies to individual cases. Finally, this study has indicated the importance of the measurement instruments used to elicit language data: future studies may explore additional strategies for balancing the need for reliable measurement with the need to elicit natural language use.

References

- Abney, S. (1987). *The English noun phrase in its sentential aspect*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge.
- Adamson, H. D. (1988). *Variation theory and second language acquisition*. Washington, DC: Georgetown University Press.
- Adamson, H. D., & Regan, V. M. (1991). The acquisition of community speech norms by Asian immigrants learning English as a second language: A preliminary study. *Studies in Second Language Acquisition, 13*, 1-22.
- Andersen, R. (1982). Determining the linguistic attributes of language attrition. In R. Lambert & B. Freed (Eds.), *The loss of language skills* (pp. 83-118). Rowley, MA: Newbury House.
- Andersen, R. (1990a). Models, processes, principles and strategies: second language acquisition inside and outside the classroom. In B. VanPatten & J. Lee (Eds.), *Second language acquisition/Foreign language learning* (pp. 45-68). Philadelphia: Multilingual Matters.
- Andersen, R. (Ed.). (1990b). Universals [Special issue]. *Studies in Second Language Acquisition, 12*.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bailey, N., Madden, C., & Krashen, S. (1974). Is there a "natural sequence" in adult second language learning? *Language Learning, 24*, 235-243.
- Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173-218). Cambridge: Cambridge University Press.
- Beck, M., & Eubank, L. (1991). Acquisition theory and experimental design: A critique of Tomasello and Herron. *Studies in Second Language Acquisition, 13*, 73-76.
- Beretta, A. (1991). Theory construction in SLA: Complementarity and opposition. *Studies in Second Language Acquisition, 13*, 493-511.
- Bialystok, E. (1988). Psycholinguistic dimensions of second language proficiency. In W. Rutherford & M. Sharwood-Smith (Eds.), *Grammar and second language teaching* (pp. 31-50). Rowley, MA: Newbury House.
- Bickerton, D. (1981). *Roots of language*. Ann Arbor: Karoma Publishers.
- Bissert, N. (1979). *Education, class, and language ideology*. London: Routledge & Kegan Paul.
- Borg, W., & Gall, M. (1989). *Educational research: An introduction* (5th ed.). New York: Longman.
- Breen, M. (1987). Contemporary paradigms in syllabus design, Parts 1 and 2. *Language Teaching, 20*, 81-92, 157-174.

- Brinton, D., Snow, M. A., & Wesche, M. B. (1989). *Content-based second language instruction*. Philadelphia: Newbury House.
- Bromley, D. (1986). *The case-study method in psychology and related disciplines*. Chichester: John Wiley & Sons.
- Brown, H. D. (1987). *Principles of language learning and teaching*. (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Brown, J. D. (1988). *Understanding research in second language learning: A teacher's guide to statistics and research design*. Cambridge: Cambridge University Press.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Brown, G., & Yule, G. (1983a). *Discourse analysis*. Cambridge: Cambridge University Press.
- Brown, G., & Yule, G. (1983b). *Teaching the spoken language: An approach based on the analysis of conversational English*. Cambridge: Cambridge University Press.
- Cadierno, T. (1992). *Explicit instruction in grammar: A comparison of input based and output based instruction in second language acquisition*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition* 15, 357-386.
- Carroll, S., Roberge, Y., & Swain, M. (1992). The role of feedback in adult second language acquisition: Error correction and morphological generalizations. *Applied Psycholinguistics*, 13, 173-198.
- Cedergren, H., & Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, 50, 333-355.
- Celce-Murcia, M. (1991). Grammar pedagogy in second and foreign language teaching. *TESOL Quarterly*, 25, 459-480.
- Celce-Murcia, M., & Larsen-Freeman, D. (1983). *The grammar book*. Cambridge, MA: Newbury house.
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. Li (Ed.), *Subject and topic* (pp. 25-56). New York: Academic Press.
- Chafe, W. (1980). *The pear stories: Cognitive, cultural and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
- Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 21-51). Amsterdam: John Benjamins.
- Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning*. Cambridge: Cambridge University Press.

- Chaudron, C., & Parker, K. (1990). Discourse markedness and structural markedness: The acquisition of English noun phrases. *Studies in Second Language Acquisition*, 12, 43-64.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1981). Principles and parameters in syntactic theory. In N. Hornstein & D. Lightfoot (Eds.), *Explanations in linguistics: The logical problem of language acquisition* (pp. 32-75). New York: Longman.
- Chomsky, N., & Lasnik, H. (1993). Principles and parameters theory. In J. Jacobs, A. von Stechow, W. Sternfeld, & T. Vennemann (Eds.), *Syntax: An international handbook of contemporary research*. Berlin: Walter de Gruyter.
- Cohen, A. (1975). Forgetting a second language. *Language Learning*, 25, 127-138.
- Collier, V. (1992). The Canadian bilingual immersion debate: A synthesis of research findings. *Studies in Second Language Acquisition*, 14, 87-97.
- Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, 15, 161-169.
- Crookes, G. (1988). *Planning, monitoring, and second language development: A review* (Tech. Rep. No. 6). Honolulu: University of Hawaii at Manoa, Social Science Research Institute, Center for Second Language Classroom Research.
- Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition*, 11, 367-383.
- Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written Communication*, 7, 482-511.
- Cumming, A., & Mellow, J. D. (1996). An investigation into the validity of written indicators of second language proficiency. In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 72-93). Clevedon, Avon: Multilingual Matters.
- Cummins, J. (1987). Immersion programs: Current issues and future directions. In L. L. Stewin & S. J. H. McCann (Eds.), *Contemporary educational issues: The Canadian mosaic* (pp. 192-206). Toronto: Copp Clark Pitman Ltd.
- Day, E., & Shapson, S. (1991). Integrating formal and functional approaches to language teaching in French immersion: An experimental study. *Language Learning*, 41, 25-58.
- de Bot, K., & Weltens, B. (1992, March). *A critical survey of research methodology in language attrition studies*. Paper presented at the conference of the American Association of Applied Linguistics, Seattle, WA.
- Delpit, L. (1986). Skills and other dilemmas of a progressive black educator. *Harvard Educational Review*, 58, 280-298.
- Delpit, L. (1988). The silenced dialogue: Power and pedagogy in educating other people's children. *Harvard Educational Review*, 56, 379-385.

- Doughty, C. (1988). *The effect of instruction on the acquisition of relativization in English as a Second Language*. Unpublished doctoral dissertation, University of Pennsylvania.
- Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition*, 13, 431-469.
- Duff, P. (1985). Syntacticization of topic in Japanese and Mandarin students' English: A test of Rutherford's model. Unpublished master's thesis, University of Hawaii.
- Eckman, F., L. Bell, & D. Nelson. (1988). On the generalization of relative clause instruction in the acquisition of English as a second language. *Applied Linguistics*, 9, 1-20.
- Eisenstein, M., & Starbuck, R. (1989). The effect of emotional investment on L2 production. In S. Gass, C. Madden, D. Preston, & L. Selinker (Eds.), *Variation in second language acquisition, Volume II: Psycholinguistic issues* (pp. 125-137). Clevedon: Multilingual Matters.
- Ellis, R. (1984). Can syntax be taught? A study of the effects of formal instruction on the acquisition of WH questions by children. *Applied Linguistics*, 5, 138-155.
- Ellis, R. (1985). Sources of variability in interlanguage. *Applied Linguistics*, 6, 118-131.
- Ellis, R. (1988). The role of practice in classroom learning. *AILA Review*, 5, 20-39.
- Ellis, R. (1989). Sources of intra-learner variability in language use and their relationship to second language acquisition. In S. Gass, C. Madden, D. Preston, & L. Selinker (Eds.), *Variation in second language acquisition: Vol. 2. Psycholinguistic issues* (pp. 22-45). Clevedon, UK: Multilingual Matters.
- Ellis, R. (1990a). *Instructed second language acquisition: Learning in the classroom*. Oxford: Basil Blackwell.
- Ellis, R. (1990b). Are classroom and naturalistic acquisition the same? A study of the classroom acquisition of German word order rules. *Studies in Second Language Acquisition*, 11, 305-328.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, R., & Roberts, C. (1987). Two approaches for investigating second language acquisition. In R. Ellis (Ed.), *Second language acquisition in context* (pp. 179-194). Englewood Cliffs, NJ: Prentice Hall.
- Fables of Aesop*. (1954). (S. A. Handford, Trans.). New York: Penguin.
- Fakhri, A. (1989). Variation in the use of referential forms within the context of foreign language loss. In S. Gass, C. Madden, D. Preston, & L. Selinker (Eds.), *Variation in second language acquisition, Volume II: Psycholinguistic issues* (pp. 189-201). Clevedon: Multilingual Matters.
- Fasold, R. (1990). *The sociolinguistics of language*. Oxford: Basil Blackwell.
- Finegan, E. (1994). *Language: Its structure and use* (2nd. ed.). Fort Worth, TX: Harcourt Brace College Publishers.

- Fukui, N., & Speas, M. (1986). Specifiers and projection. *MIT Working Papers in Linguistics*, 8, 128-172.
- Gardner, R., Lalonde, R., & MacPherson, J. (1985). Social factors in second language attrition. *Language Learning*, 35, 519-540.
- Gary, J. D., & Gary, N. (1981). Talking may be dangerous to your linguistic health: The case for a much greater emphasis on listening comprehension in foreign language instruction. *International Review of Applied Linguistics*, 19, 1-14.
- Gass, S. (1982). From theory to practice. In M. Hines & W. Rutherford (Eds.), *On TESOL '81* (pp. 129 - 139). Washington, DC: TESOL.
- Gass, S., & Selinker, L. (1994). *Second language acquisition: An introductory course*. Hillsdale, NJ: Lawrence Erlbaum.
- Ghiselli, E., Campbell, J., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. New York: W. H. Freeman and Company.
- Givon, T. (1984). Universals of discourse structure and second language acquisition. In W. Rutherford (Ed.), *Language universals and second language acquisition* (pp. 109-136). Amsterdam: John Benjamins.
- Glass, G. V. (Speaker). (1980). *Quasi-experimental research methods in education: Time series experiments*. (Cassette recording and study guide) Washington, DC: American Educational Research Association.
- Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology* (2nd. Ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Goodman, K., Smith, E. B., Meredith, R., & Goodman, Y. (1987). *Language and thinking in school: A whole-language curriculum*. New York: Richard C. Owen.
- Graman, T. (1988). Education for humanization: Applying Paulo Freire's pedagogy to learning a second language. *Harvard Educational Review*, 58, 433-448.
- Gregg, K. (1984). Krashen's monitor and Occam's razor. *Applied Linguistics*, 5, 79-100.
- Gregg, K. (1989). Second language acquisition theory: The case for a generative perspective. In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 15-40). Cambridge: Cambridge University Press.
- Gregg, K. (1990). The variable competence model of second language acquisition and why it isn't. *Applied Linguistics*, 11, 364-383.
- Gregg, K. (1993). Taking explanation seriously; or, let a couple of flowers bloom. *Applied Linguistics*, 14, 276-294.
- Grondin, N. (1992). *Functional projections in child second language acquisition of French*. Unpublished master's thesis, McGill University, Montreal.
- Guilfoyle, E., & Noonan, M. (1992). Functional categories and language acquisition. *The Canadian Journal of Linguistics*, 37, 241-272.

- Hammerly, H. (1992). The need for directed language learning in the FL classroom: A response to Collier. *Studies in Second Language Acquisition*, 14, 215-216.
- Handscombe, J. (1990). The complementary roles of researchers and practitioners in second language education. In B. Harley, P. Allen, J. Cummins, & M. Swain (Eds.), *The development of second language proficiency* (pp. 181-186). Cambridge: Cambridge University Press.
- Hangsen, R. D. (1991, May). Can education become a science? *Phi Delta Kappan*, 689- 694.
- Harley, B. (1989a). Functional grammar in French immersion: A classroom experiment. *Applied Linguistics*, 10, 331-358.
- Harley, B. (1989b). Effects of instruction on SLA: Issues and evidence. *Annual Review of Applied Linguistics* (1988), 9, 165-178.
- Harley, B., Cummins, J., Swain, M., & Allen, P. (1990). The nature of language proficiency. In B. Harley, P. Allen, J. Cummins, & M. Swain (Eds.), *The development of second language proficiency* (pp. 7-25). Cambridge: Cambridge University Press.
- Hatch, E. (1983). *Psycholinguistics: A second language perspective*. Rowley, MA: Newbury House.
- Hatch, E. (1992). *Discourse and language education*. Cambridge: Cambridge University Press.
- Hatch, E., & Lazarson, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House.
- Hedgcock, J. (1991). *Foreign language retention and attrition: A study of maintenance patterns among adult learners of Spanish*. Unpublished doctoral dissertation, University of Southern California, Los Angeles.
- Hedgcock, J. (1994, March). *Memory and transfer effects in L2 attrition: An analysis of forms that 'stick' and forms that don't*. Paper presented at the meeting of the American Association for Applied Linguistics, Baltimore, MD.
- Higgs, T., & Clifford, R. (1982). The push toward communication. In T. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher* (pp. 57-79). Skokie, IL: National Textbook Company.
- Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition Vol. 1: Foundations* (pp. 77-109). Cambridge, MA: The MIT Press.
- Hubbard, P., Jones, H., Thornton, B., & Wheeler, R. (1983). *A training course for TEFL*. Oxford: Oxford University Press.
- Huebner, T. (1979). Order-of-acquisition vs. dynamic paradigm: A comparison of method in interlanguage research. *TESOL Quarterly*, 13, 21-28.
- Huebner, T. (1983a). Linguistic systems and linguistic change in an interlanguage. *Studies in Second Language Acquisition*, 6, 33-53.

- Huebner, T. (1983b). *A longitudinal analysis of the acquisition of English*. Ann Arbor: Karoma.
- Huebner, T., Perdue, C., & Carroll, M. (1988). The acquisition of English. In W. Klein & C. Perdue (Eds.), *Utterance structure* (pp. 36-136). Final Report, Vol. VI of Second Language Acquisition by Adult Immigrants: An Additional Activity of the European Science Foundation.
- Hulstijn, J., & Hulstijn, W. (1984). Grammatical errors as a function of processing constraints and explicit knowledge. *Language Learning*, 34, 23-44.
- Ishiyama, F. I. (1991). A Japanese reframing technique for brief social anxiety treatment: An exploratory study of cognitive and therapeutic effects of Morita therapy. *Journal of Cognitive Psychotherapy: An International Quarterly*, 5, 55-70.
- Ingram, D. (1989). *First language acquisition: Method, description, and explanation*. Cambridge: Cambridge University Press.
- Johnson, D. (1992). *Approaches to research in second language learning*. New York: Longman.
- Johnson, R. K. (1989). A decision-making framework for the coherent language curriculum. In R. K. Johnson (Ed.), *The second language curriculum* (pp. 1-23). Cambridge: Cambridge University Press.
- Kadia, K. (1988). The effect of formal instruction on monitored and on spontaneous naturalistic interlanguage performance: A case study. *TESOL Quarterly*, 22, 509-515.
- Kaufman, D. (1992, March). *The process of L1 verb-system disintegration followed by reconstructed forms in Hebrew speaking children living in the U.S.* Paper presented at the conference of the American Association of Applied Linguistics, Seattle, WA.
- Kazdin, A. (1982). *Single-case research designs: Methods for clinical and applied settings*. New York: Oxford University Press.
- Kennedy, E. (1988). *The oral interaction of native speakers and non-native speakers in a multicultural preschool: A comparison between freeplay and contrived NS/NNS dyads*. Unpublished master's thesis, University of British Columbia, Vancouver, Canada.
- Klein, W. (1990). A theory of language acquisition is not so easy. *Studies in Second Language Acquisition*, 12, 219-231.
- Klein, W., & Perdue, C. (Eds.). (1988). *Utterance structure*. Final Report, Vol. VI of Second Language Acquisition by Adult Immigrants: An Additional Activity of the European Science Foundation.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. New York: Pergamon Press.
- Krashen, S. (1992). Under what circumstances, if any, should formal grammar instruction take place? *TESOL Quarterly*, 26, 409-411.
- Krashen, S. (1993). The effect of formal grammar teaching: Still peripheral. *TESOL Quarterly*, 27, 722-725.

- Kratochwill, T. (Ed.). (1978). *Single subject research: Strategies for evaluating change*. New York: Academic Press.
- Krishef, C. (1991). *Fundamental approaches to single subject design and analysis*. Malabar, FL: Krieger.
- Kumpf, L. (1992, April). *New information and full noun distribution in second language discourse*. Paper presented at the Second Language Research Forum, East Lansing, Michigan.
- Labov, W. (1970). The study of language in its social contexts. *Studium Generale*, 23, 30-87.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Lalande, J. (1982). Reducing composition errors: An experiment. *Modern Language Journal*, 66, 140-149.
- Lambert, R. D., & Moore, S. J. (1986). Problem areas in the study of language attrition. In B. Weltens, K. de Bot, & T. van Els (Eds.), *Language attrition in progress* (pp. 177-186). Dordrecht: Foris.
- Lantolf, J., & Ahmed, M. (1989). Psycholinguistic perspectives on interlanguage variation: A Vygotskyian analysis. In S. Gass, C. Madden, D. Preston, & L. Selinker (Eds.), *Variation in second language acquisition, Volume II: Psycholinguistic issues* (pp. 93-108). Clevedon: Multilingual Matters.
- Lantolf, J., & Frawley, W. (1988). Proficiency: Understanding the construct. *Studies in Second Language Acquisition*, 10, 181-195.
- Larsen-Freeman, D. (1986). *Techniques and principles in language teaching*. Oxford: Oxford University Press.
- Larsen-Freeman, D. (1991). Second language acquisition research: Staking out the territory. *TESOL Quarterly*, 25, 315-350.
- Larsen-Freeman, D., & Long, M. (1991). *An introduction to second language acquisition research*. London: Longman.
- Lazarton, A. (1991). Power, effect size, and second language research: A researcher comments. *TESOL Quarterly*, 25, 759-762.
- Lewontin, R. C., Rose, S., & Kamin, L. J. (1984). *Not in our genes: Biology, ideology and human nature*. New York: Pantheon.
- Lightbown, P. & Spada, N. (1990). Focus-on-form and corrective feedback in communicative language teaching: Effects on second language learning. *Studies in Second Language Acquisition*, 12, 429-448.
- Lightbown, P. & Spada, N. (1993). *How languages are learned*. Oxford: Oxford University Press.
- Lightbown, P., N. Spada, & R. Wallace. (1980). Some effects of instruction on child and adolescent ESL learners. In R. Scarcella & S. Krashen (Eds.), *Research in second language acquisition* (pp. 162-171). Rowley, MA: Newbury House.

- Long, M. (1983). Does second language instruction make a difference? A review of research. *TESOL Quarterly*, 17, 359-382.
- Long, M. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives* (pp. 115-141). New York: Newbury House.
- Long, M. (1993). Assessment strategies for second language acquisition theories. *Applied Linguistics*, 14, 225-249.
- Long, M. & Sato, C. (1984). Methodological issues in interlanguage studies: An interactionist perspective. In A. Davies, C. Cramer, & A. Howatt (Eds.), *Interlanguage* (pp. 253-279). Edinburgh: Edinburgh University Press.
- Maratsos, M. (1974). Preschool children's use of definite and indefinite articles. *Child Development*, 45, 446-455.
- Maratsos, M. (1979). Learning how and when to use pronouns and determiners. In P. Fletcher & M. Garman (Eds.), *Language acquisition: Studies in first language development* (pp. 225-240). Cambridge: Cambridge University Press.
- Master, P. (1987). *A cross-linguistic interlanguage analysis of the acquisition of the English article system*. Unpublished doctoral dissertation, University of California at Los Angeles.
- Master, P. (1990). Teaching the English articles as a binary system. *TESOL Quarterly*, 24, 461-478.
- Master, P. (1994). The effect of systematic instruction on learning the English article system. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 229-249). Cambridge: Cambridge University Press.
- Masterson, D. (1989). Functional categories and parameter setting in L2 acquisition. *Korean Linguistics*, 6, 179-192.
- McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.
- McCleary, R., & Hay, R. (1980). *Applied time series analysis for the social sciences*. Beverly Hills, CA: Sage.
- McKenna, F. R. (1976, February). Piaget's complaint - and mine: Why is there no science of education? *Phi Delta Kappan*, 405-409.
- McLaughlin, B. (1987). *Theories of second-language learning*. London: Edward Arnold.
- McLaughlin, B. (1990). Restructuring. *Applied Linguistics*, 11, 113-128.
- Mellow, J. D. (1988). On triggers: A parameter-setting approach to the acquisition of Cree, a free word order language. *McGill Working Papers in Linguistics*, 5, 2, 97-127.
- Mellow, J. D. (1992, April). *Towards a theory of second language transition: Implications from constrained studies of instruction and attrition*. Paper presented at the Second Language Research Forum, East Lansing, Michigan.

- Mellow, J. D. (1996). *A preliminary study of the acquisition of functional categories by Japanese ESL learners*. Unpublished manuscript, University of British Columbia.
- Mellow, J. D. (in press). On the primacy of theory in applied studies: A critique of Pienemann and Johnston (1987). *Second Language Research*.
- Mellow, J. D., & Cumming, A. (1994). Concord in interlanguage: Efficiency or priming? *Applied Linguistics*, 15, 442-473.
- Mellow, J. D., & Fuller, J. (1995, September). *The effect of instruction on the ESL acquisition of noun phrase reference*. Paper presented at the Second Language Research Forum, Ithaca, NY.
- Mellow, J. D., Reeder, K., & Forster, E. (in press). Using the time-series design to investigate the effects of pedagogic intervention on SLA. *Studies in Second Language Acquisition*, 18.
- Mohan, B. (1986). *Language and content*. Reading, MA: Addison-Wesley.
- Nunan, D. (1991). Methods in second language classroom-oriented research: A critical review. *Studies in Second Language Acquisition*, 13, 249-274.
- O'Malley, J. & Chamot, A. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
- O'Malley, J.M., Chamot, A. U., Stewner-Manzanares, G., Russo, R., & Kupper, L. (1985). Learning strategy applications with students of English as a second language. *TESOL Quarterly*, 19, 285-296.
- Owens, R. (1984). *Language development: An introduction*. Columbus, OH: Charles E. Merrill.
- Oxford, R. (1982). Research on language loss: A review with implications for foreign language teaching. *Modern Language Journal*, 66, 160-169.
- Parrish, B. (1987). A new look at methodologies in the study of article acquisition for learners of ESL. *Language Learning*, 37, 361-383.
- Parsonson, B., & Baer, D. (1978). The analysis and presentation of graphic data. In T. Kratochwill (Ed.), *Single subject research: Strategies for evaluating change* (pp. 101-165). New York: Academic Press.
- Pfaff, C. (1987). Functional approaches to interlanguage. In C. Pfaff (Ed.), *First and second language acquisition processes* (pp. 81-102). New York: Newbury House.
- Pica, T. (1983a). Methods of morpheme quantification: Their effect on the interpretation of second language data. *Studies in Second Language Acquisition*, 6, 69-78.
- Pica, T. (1983b). The article in American English: What the textbooks don't tell us. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 222-233). Rowley, MA.: Newbury House.
- Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communication tasks for second language instruction and research. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 9-34). Philadelphia: Multilingual Matters.

- Pienemann, M. (1984). Psychological constraints on the teachability of language. *Studies in Second Language Acquisition*, 6, 109-135.
- Pienemann, M. (1985). Learnability and syllabus construction. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition*. (pp. 23-76). Clevedon, UK: Multilingual Matters.
- Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, 10, 52-79.
- Preston, R. (1989). *Sociolinguistics and second language acquisition*. Oxford: Basil Blackwell.
- Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223-255). New York: Academic Press.
- Prince, E. (1992). The ZPG letter: Subjects, definiteness, and information-status. In W. C. Mann & S. A. Thompson (Eds.), *Discourse description: Diverse linguistic analyses of a fundraising text* (pp. 295-325). Amsterdam: John Benjamins.
- Quirk, R., & Greenbaum, S. (1973). *A university grammar of English*. Burnt Mill, UK: Longman.
- Radford, A. (1981). *Transformational syntax: A student's guide to Chomsky's Extended Standard Theory*. Cambridge: Cambridge University Press.
- Radford, A. (1990). *Syntactic theory and the acquisition of English syntax*. Oxford: Basil Blackwell.
- Rampton, B. (1987). Stylistic variability and not speaking 'normal' English: Some post-Labovian approaches and their implications for the study of interlanguage. In R. Ellis (Ed.), *Second language acquisition in context* (pp. 47-58). Englewood Cliffs, NJ: Prentice Hall.
- Reeder, K., & Shapiro, J. (1993). Relationships between early literate experience and knowledge and children's linguistic pragmatic strategies. *Journal of Pragmatics*, 19, 1-22.
- Richards, J., & Rodgers, T. (1986). *Approaches and methods in language teaching: A description and analysis*. Cambridge: Cambridge University Press.
- Ritter, E. (1992) Cross-linguistic evidence for number phrase. *Canadian Journal of Linguistics*, 37, 197-218.
- Ross, S., & Berwick, R. (1991). The acquisition of English affixes through general and specific instructional strategies. *JALT Journal*, 13, 131-142.
- Runkel, P. J. (1990). *Casting nets and testing specimens: Two grand methods of psychology*. New York: Praeger.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 17-46.
- Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language. In R. Day (Ed.), *Talking to learn: Conversation and second language acquisition* (pp. 237-326). Rowley, MA: Newbury House.

- Schulz, C. M. (1973). *The Charlie Brown dictionary*. New York: World Publishing.
- Schumann, J. (1978). Second language acquisition: The pidginization hypothesis. In E. Hatch (Ed.), *Second language acquisition: A book of readings* (pp. 256-271). Rowley, MA: Newbury House.
- Seliger, H. & Shohamy, E. (1989). *Second language research methods*. Oxford: Oxford University Press.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 209-31.
- Shibatani, M. (1990). *The languages of Japan*. Cambridge: Cambridge University Press.
- Sharwood Smith, M. (1981). Consciousness-raising and the second language learner. *Applied Linguistics*, 11, 159-168.
- Sharwood Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research*, 7, 118-32.
- Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, 15, 165-179.
- Slobin, D. (1985). *The cross-linguistic study of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum.
- Skehan, P. (1987). Variability and language testing. In R. Ellis (Ed.), *Second language acquisition in context* (pp. 195-206). Englewood Cliffs, NJ: Prentice Hall.
- Spada, N. & Lightbown, P. (1993). Instruction and the development of questions in L2 classrooms. *Studies in Second Language Acquisition*, 15, 205-224.
- Spolsky, B. (1989). *Conditions for second language learning*. Oxford: Oxford University Press.
- Standards for educational and psychological testing*. (1985). Washington, DC: The American Psychological Association.
- Stern, H. (1983). *Fundamental concepts of language teaching*. Oxford: Oxford University Press.
- Stevens, J. (1986). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Strauss, S, (Ed.). (1982). *U-shaped behavioral growth*. New York: Academic Press.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.
- Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *Canadian Modern Language Review*, 50, 158-164.
- Swan, M. (1987). Non-systematic variability: A self-inflicted conundrum? In R. Ellis (Ed.), *Second language acquisition in context* (pp. 59-66). Englewood Cliffs, NJ: Prentice Hall.

- Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics* (2nd. ed.). New York: HarperCollins.
- Tarone, E. (1985). Variability in interlanguage use: A study of style-shifting in morphology and syntax. *Language Learning*, 35, 373-404.
- Tarone, E. (1987). Methodologies for studying variability in second language acquisition. In R. Ellis (Ed.), *Second language acquisition in context* (pp. 35-46). Englewood Cliffs, NJ.: Prentice Hall.
- Tarone, E. (1988). *Variation in interlanguage*. London: Edward Arnold.
- Tarone, E. (1993, October). *Second-language acquisition in a variationist framework*. Paper presented at the 22nd University of Wisconsin-Milwaukee Linguistics Symposium, Milwaukee.
- Tarone, E., & Parrish, B. (1988). Task-related variation in interlanguage: The case of articles. *Language Learning*, 38, 21-44.
- Tawney, J., & Gast, D. (1984). *Single subject research in special education*. Columbus, OH: Charles E. Merrill.
- Terrell, T. (1991). The role of grammar instruction in a communicative approach. *Modern Language Journal*, 75, 52-63.
- Thomas, M. (1989). The acquisition of English articles by first- and second-language learners. *Applied Psycholinguistics*, 10, 335-355.
- Tomasello, M., & Herron, C. (1988). Down the garden path: Inducing and correcting overgeneralization errors in the foreign classroom. *Applied Psycholinguistics*, 9, 237-246.
- Tomasello, M., & Herron, C. (1989). Feedback for language transfer errors: the garden path technique. *Studies in Second Language Acquisition*, 11, 385-395.
- Tomasello, M., & Herron, C. (1991). Experiments in the real world: A reply to Beck and Eubank. *Studies in Second Language Acquisition*, 13, 513-517.
- Tomlin, R. (1990). Functionalism in second language acquisition. *Studies in Second Language Acquisition*, 12, 155-177.
- Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.
- van Els, T. (1986). An overview of European research on language attrition. In B. Weltens, K. de Bot, & T. van Els (Eds.), *Language attrition in progress* (pp. 3-18). Dordrecht: Foris.
- van Ek, J. A. (1987). The threshold level. In M. H. Long & J. C. Richards (Eds.), *Methodology in TESOL: A book of readings* (pp. 78-85). New York: Newbury House.
- VanPatten, B. (1984). Processing strategies and morpheme acquisition. In F. Eckman, L. Bell, & D. Nelson (Eds.), *Universals in second language acquisition* (pp. 88-98). Rowley, MA: Newbury House.

- VanPatten, B. (1988). How juries get hung: Problems with the evidence for a focus on form in teaching. *Language Learning*, 38, 243-260.
- VanPatten, B. (1993). Grammar teaching for the acquisition-rich classroom. *Foreign Language Annals*, 26, 435-450.
- VanPatten, B. (1994). Cognitive aspects of processing in second language acquisition. In P. Hashemipour, I. Maldonado, & M. van Naerssen (Eds.), *Festschrift for Tracy David Terrell*. New York: McGraw-Hill.
- VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, 15, 225-243.
- von Stutterheim, C., & Klein, W. (1987). A concept-oriented approach to second language studies. In C. Pfaff (Ed.), *First and second language acquisition processes* (pp. 191-205). New York: Newbury House.
- Wells, G. (1986). Variation in child language. In P. Fletcher & M. Garman (Eds.), *Language acquisition: Studies in first language development* (pp. 109-139). Cambridge: Cambridge University Press.
- Weltens, B. (1987). The attrition of foreign-language skills: A literature review. *Applied Linguistics*, 8, 22-38.
- Weltens, B., & Cohen, A. (1989). Language attrition research: An introduction. *Studies in Second Language Acquisition*, 11, 227-233.
- White, L. (1987). Against comprehensible input: The input hypothesis and the development of second language competence. *Applied Linguistics*, 8, 95-110.
- White, L. (1989). *Universal grammar and second language acquisition*. Amsterdam: John Benjamins.
- White, L. (1991). Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Second Language Research*, 7, 133-161.
- White, L., Spada, N., Lightbown, P., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics*, 12, 416-432.
- Woods, A., Fletcher, P., & Hughes, A. (1986). *Statistics in language studies*. Cambridge: Cambridge University Press.
- Yamada, J., & Matsuura, N. (1982). The use of the English article among Japanese students. *RELC Journal*, 13, 50-63.
- Yin, R. (1989). *Case study research: Design and methods* (rev. ed.). Newbury Park: Sage Publications.
- Young, R. (1988). Variation and interlanguage hypothesis. *Studies in Second Language Acquisition*, 10, 281-302.
- Young, R. (1989). Ends and means: Methods for the study of interlanguage variation. In S. Gass, C. Madden, D. Preston, & L. Selinker (Eds.), *Variation in second language acquisition, Volume II: Psycholinguistic issues* (pp. 65-90). Clevedon: Multilingual Matters.

- Young, R. (1993). Functional constraints on variation in interlanguage morphology. *Applied Linguistics*, 14, 76-97.
- Young, R. (1996). Form-function relations in articles in English interlanguage. In R. Bayley & D. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 135-175). Amsterdam: John Benjamins.
- Yule, G. (1981). New, current and displaced identity reference. *Lingua*, 55, 41-52.
- Zobl, H. (1985). Grammars in search of input and intake. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 329-344). Rowley, MA: Newbury House.

Appendix A

Consent Form

THE UNIVERSITY OF BRITISH COLUMBIA



Department of Language Education
2125 Main Mall
Vancouver, B.C. Canada V6T 1Z4
Tel: (604) 822-5788
Fax: (604) 822-3154

Project Faculty Advisor:
Dr. Reeder - 822-5764

To: Tutoring Participants
From: Dean Mellow
Regarding: The Research Project - An investigation into the development of ESL ability

For my research as a university student, I am investigating the improvement in language ability that results from living in Vancouver, taking courses at UBC, and receiving additional tutoring. I am asking you to participate in 1 1/2 hour tutoring sessions, once each week from September 16th to December 5, 1992. If you are interested, the sessions may continue the following semester. The topics that we will study include: organizing and editing essays, developing appropriate word choice, and aspects of grammar that are difficult for Japanese speakers. I am requesting that you will regularly attend these sessions and that you will allow me to record your speech, make copies of your writing assignments, and have you do some in-class assignments.

When I report the results of your learning for my study, your identity will be kept confidential: I will use a codename for you.

You may refuse to participate in this project and you may withdraw at any time. Your participation in the tutoring, your refusal to participate in the tutoring, or your withdrawal from the tutoring will not affect your course grades. However, I do believe that the tutoring may help you to improve your English skills.

If you do not understand the procedures or have any questions about the project, please ask Dean for more information.

Thank you for your assistance.

Dean Mellow

Signature of Consent:

I consent to participate in this project and to allow my writing, speech and assignments to be used for the study.

Name

Date

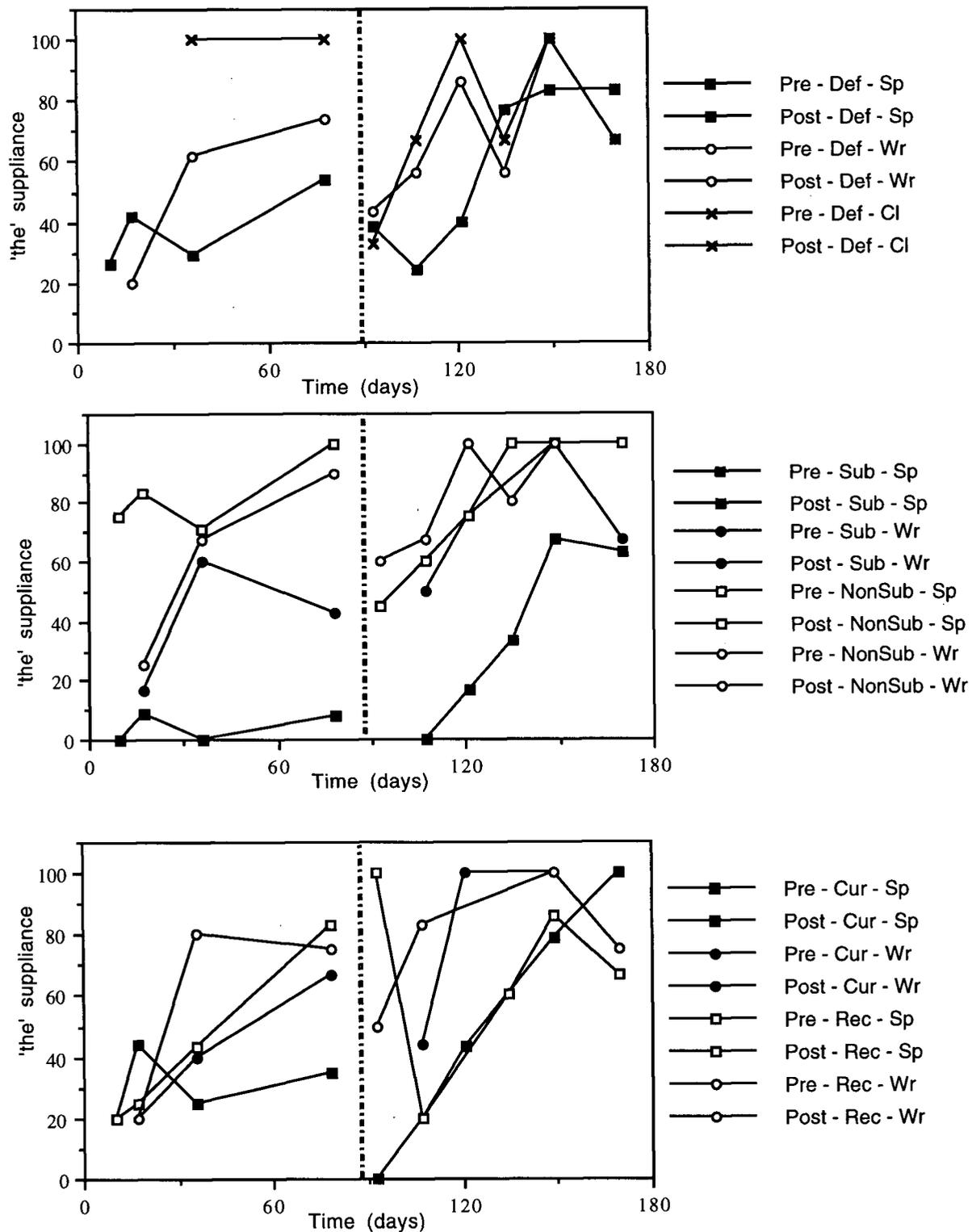
I also acknowledge that I have received a copy of this consent form for my own records.

Name

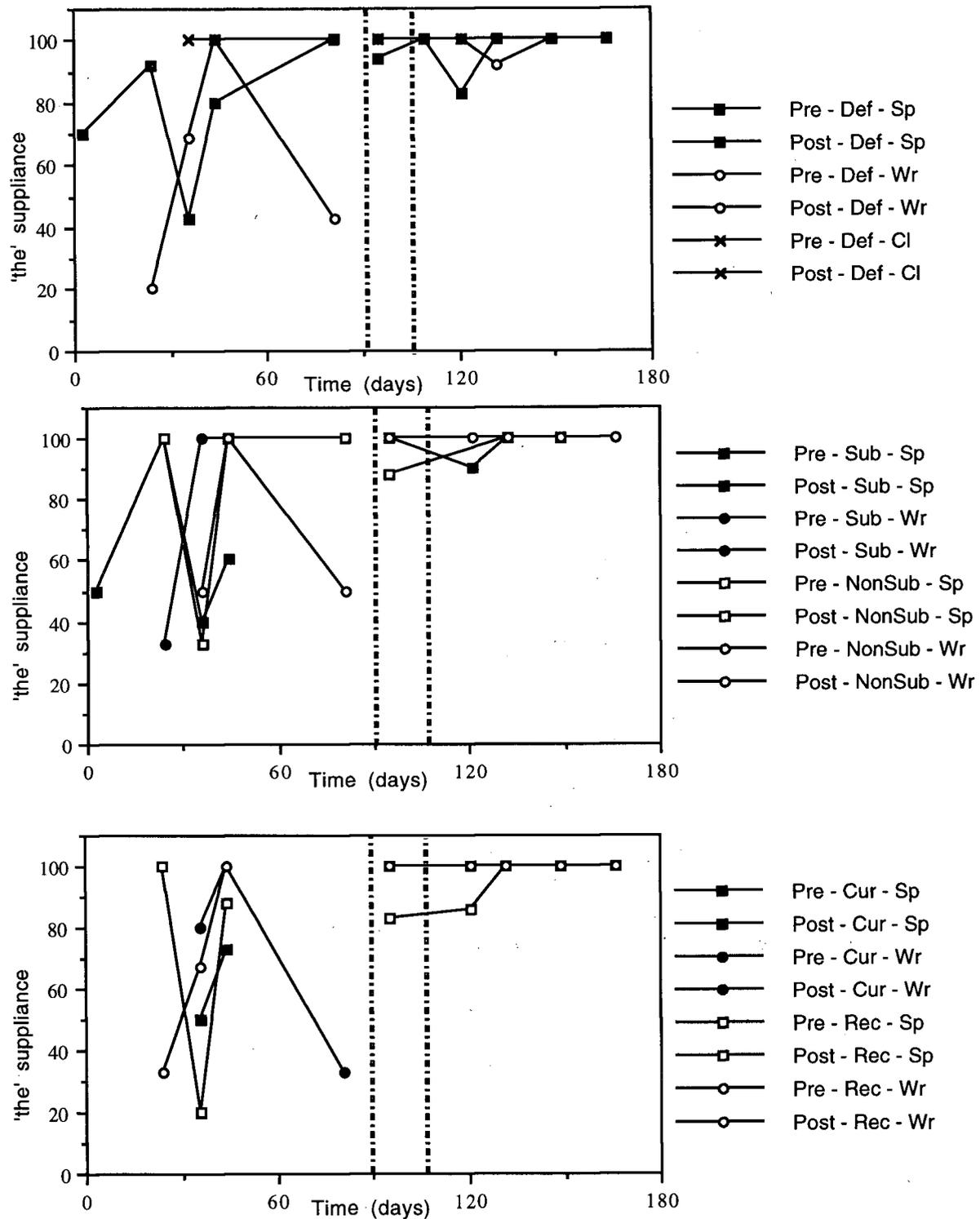
Date

Appendix B

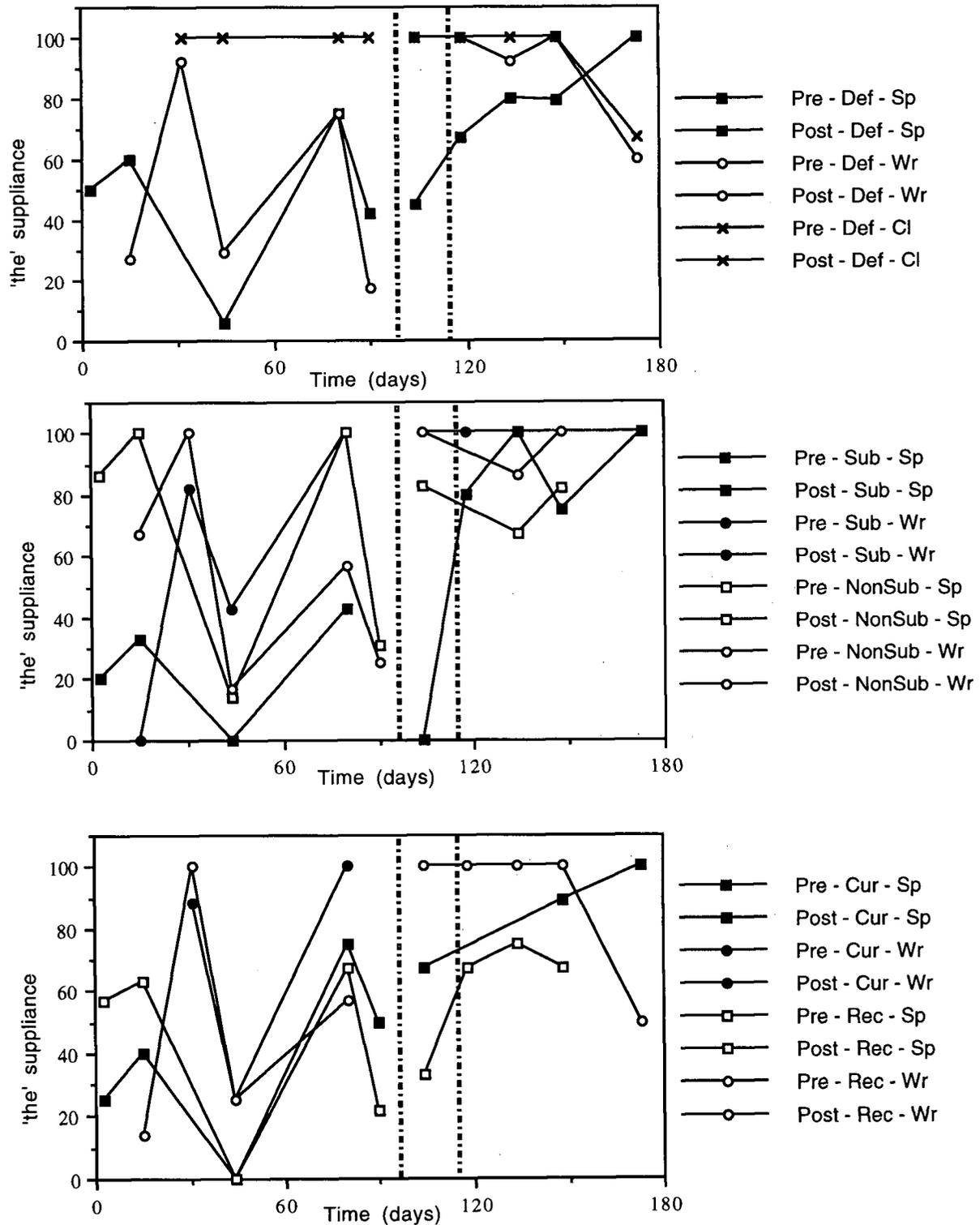
Line Graphs for All Learners for All Dependent Variables



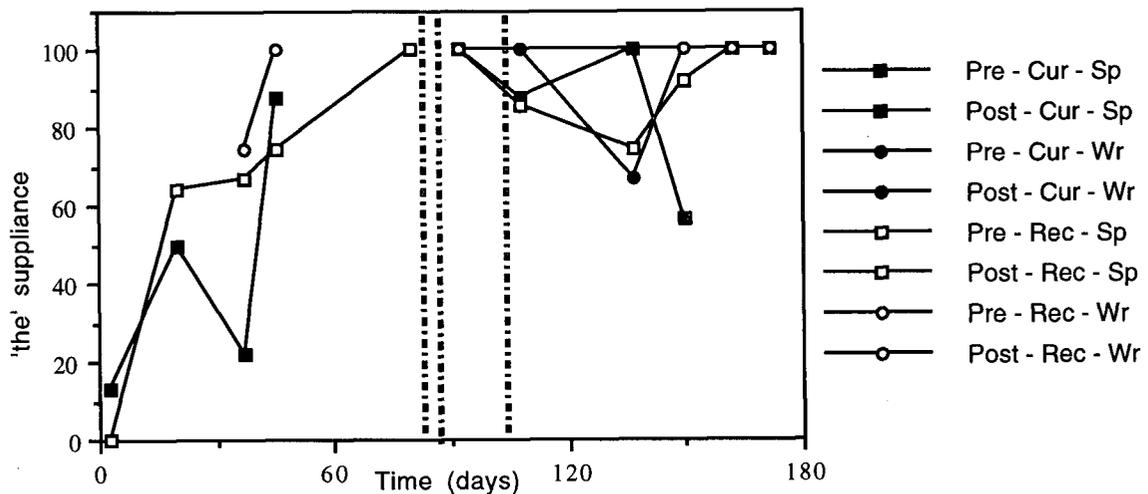
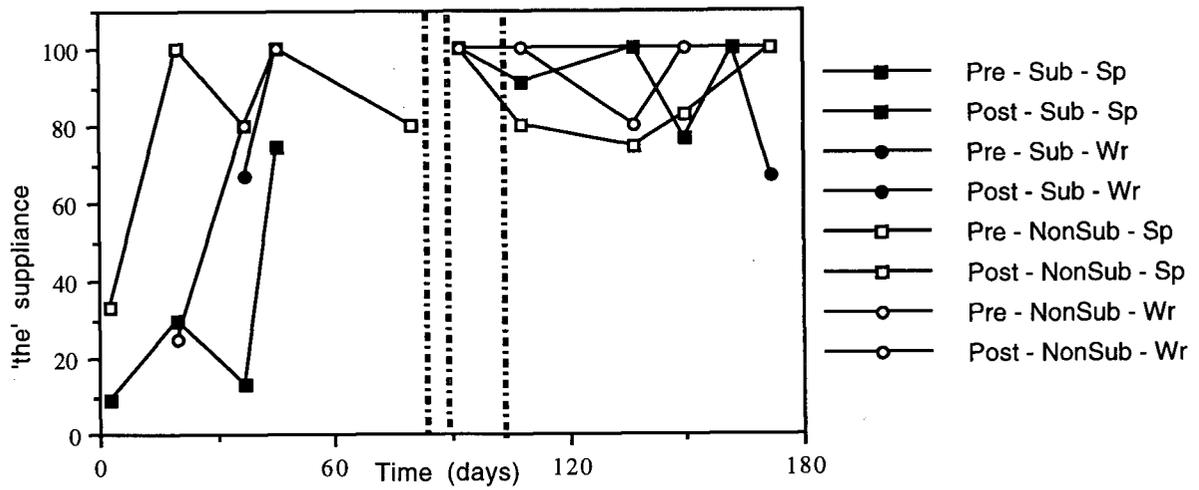
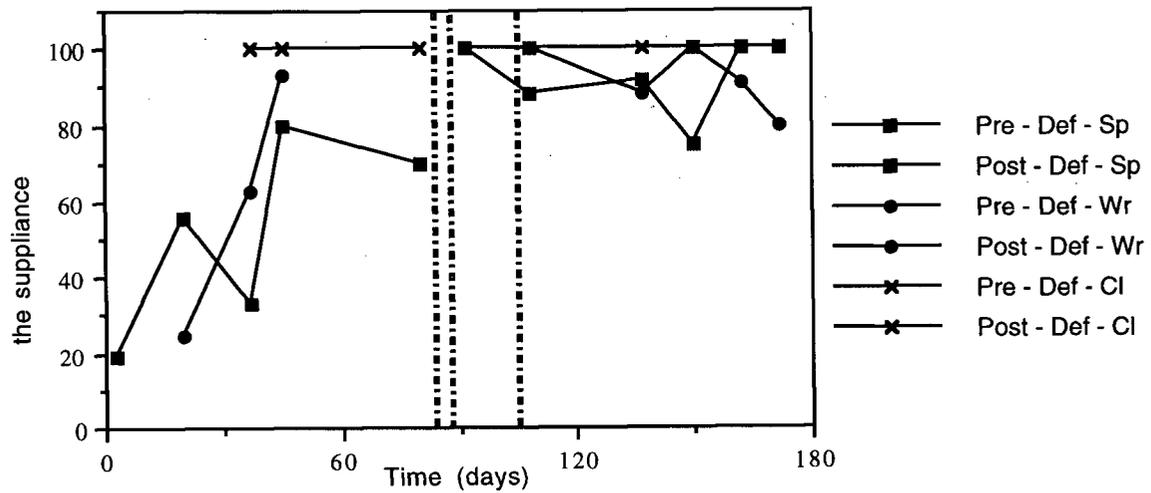
Appendix B1: Learner Tr1, Line Graphs of Developmental Course, For All Dependent Variables on All Tasks, With Lines Connecting the Data Points Within Each Phase, and With the Session of Instruction Indicated by a Vertical Line



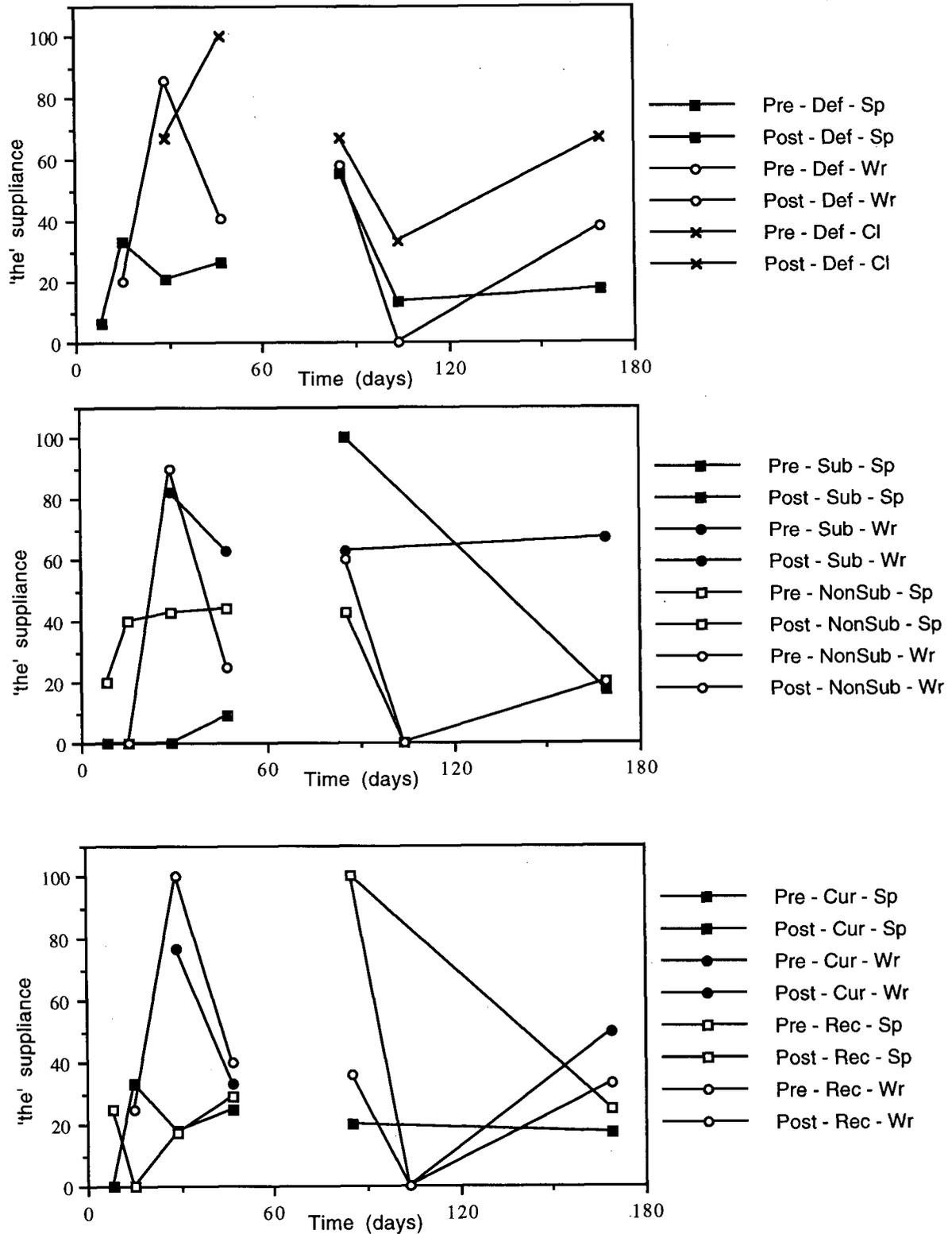
Appendix B2: Learner Tr2, Line Graphs of Developmental Course, For All Dependent Variables on All Tasks, With Lines Connecting the Data Points Within Each Phase, and With Sessions of Instruction Indicated by Vertical Lines



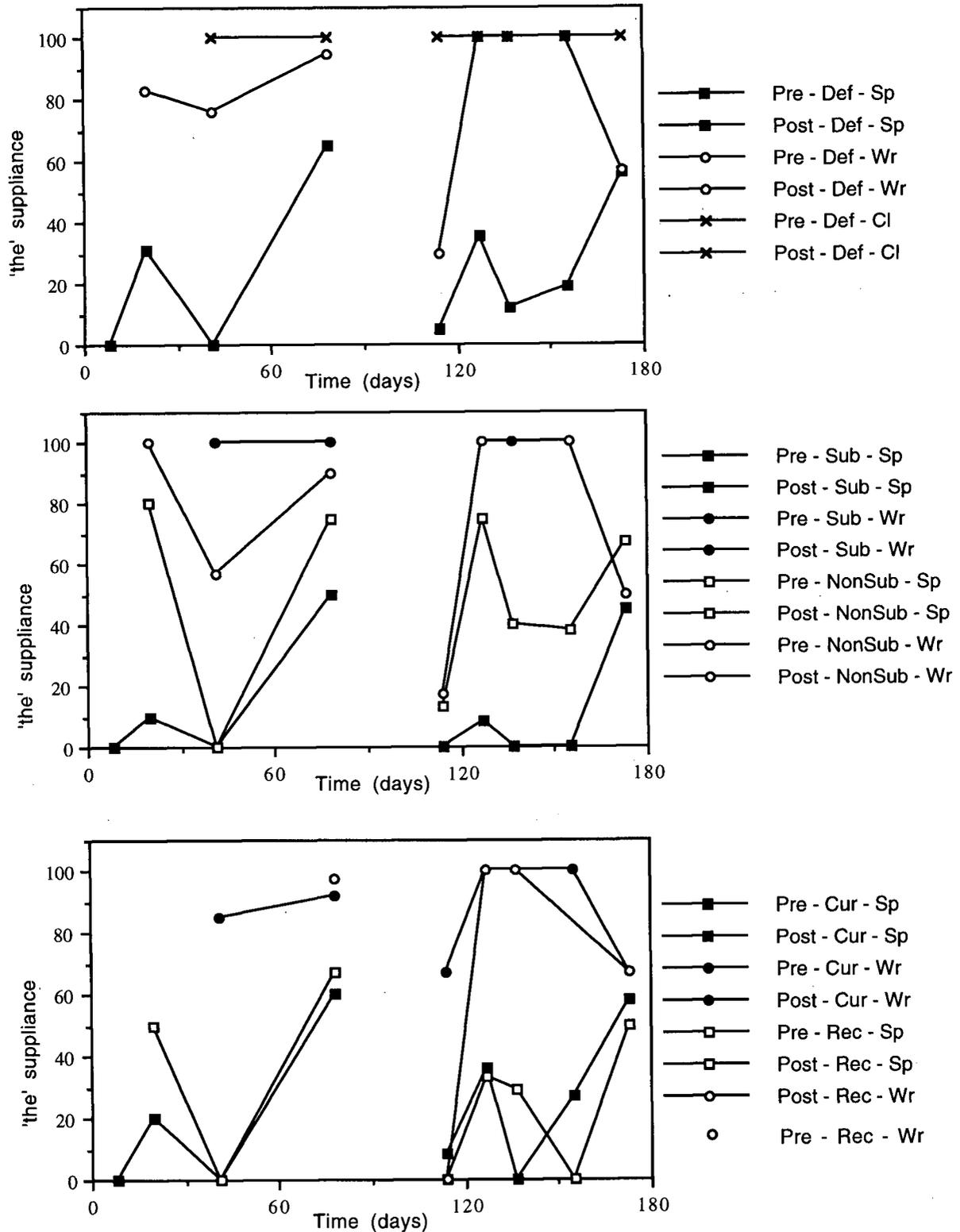
Appendix B3: Learner Tr3, Line Graphs of Developmental Course, For All Dependent Variables on All Tasks, With Lines Connecting the Data Points Within Each Phase, and With Sessions of Instruction Indicated by Vertical Lines



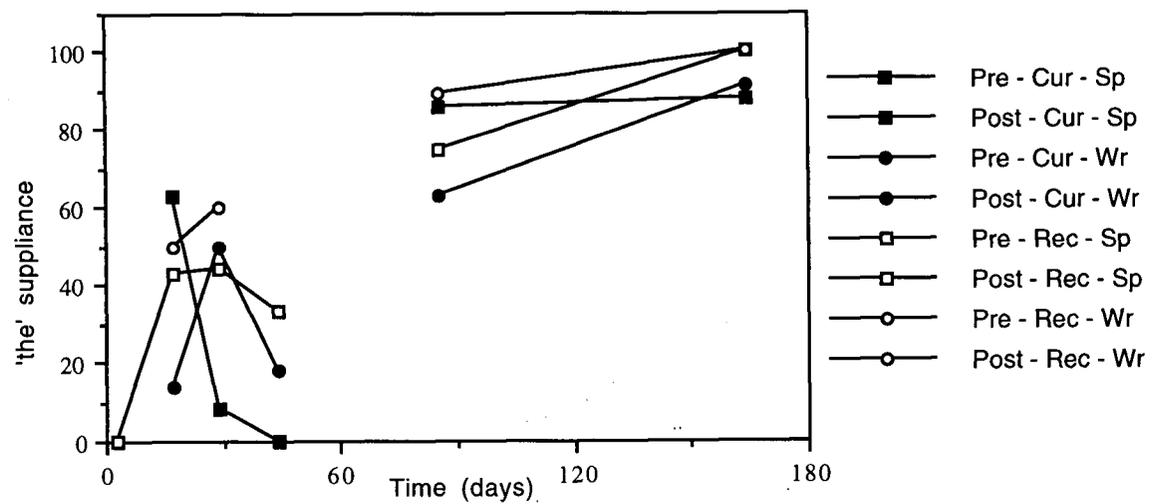
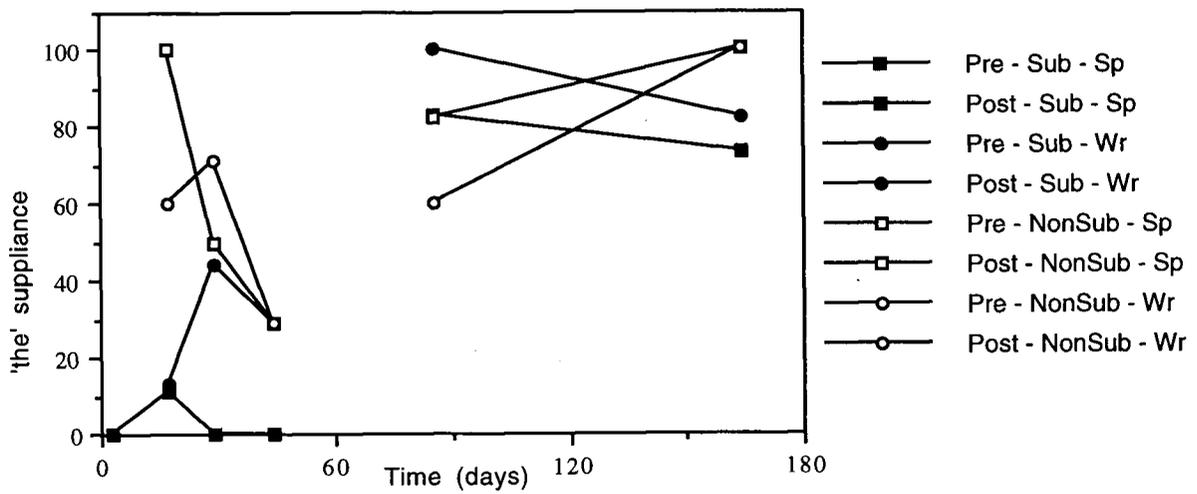
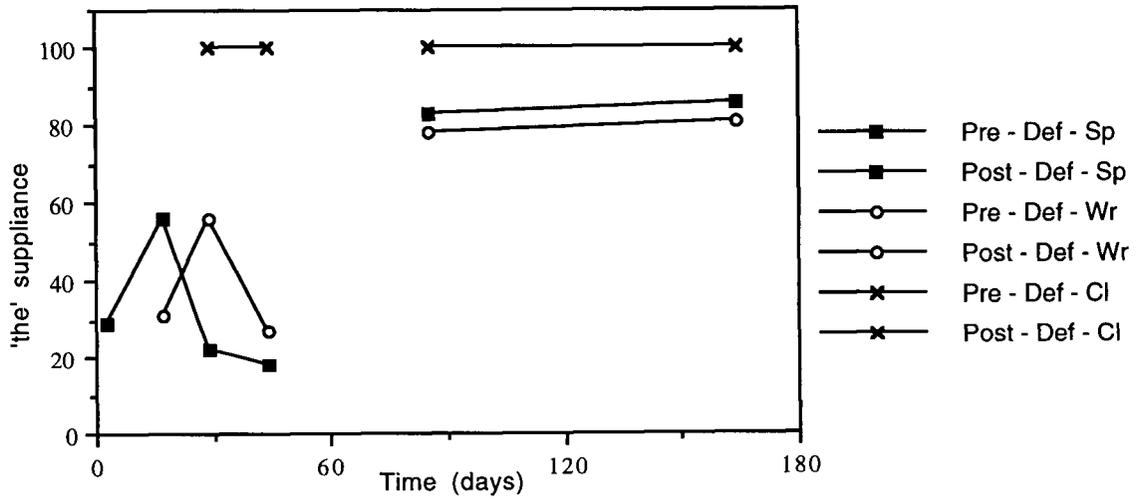
Appendix B4: Learner Tr4, Line Graphs of Developmental Course, For All Dependent Variables on All Tasks, With Lines Connecting the Data Points Within Each Phase, and With Sessions of Instruction Indicated by Vertical Lines



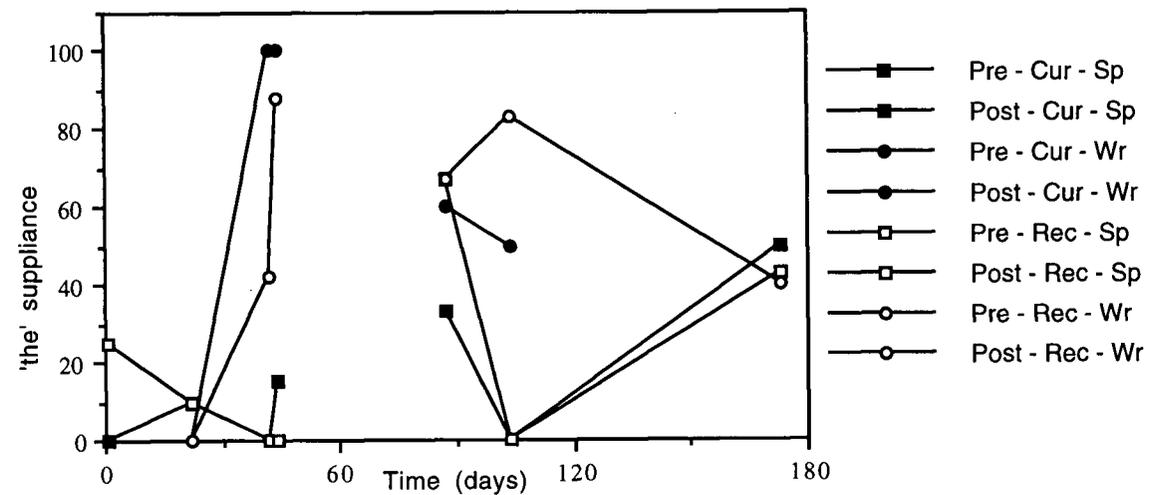
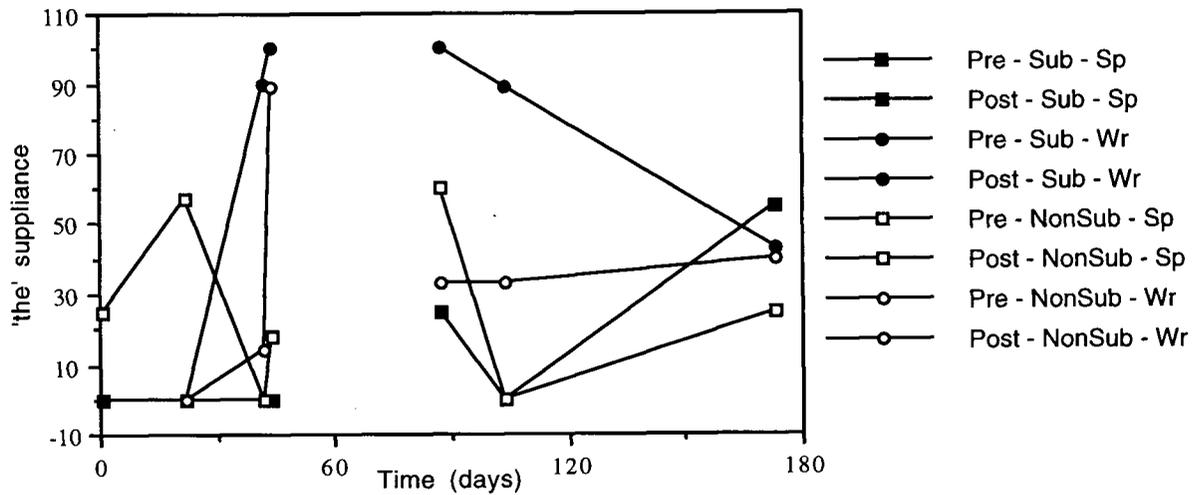
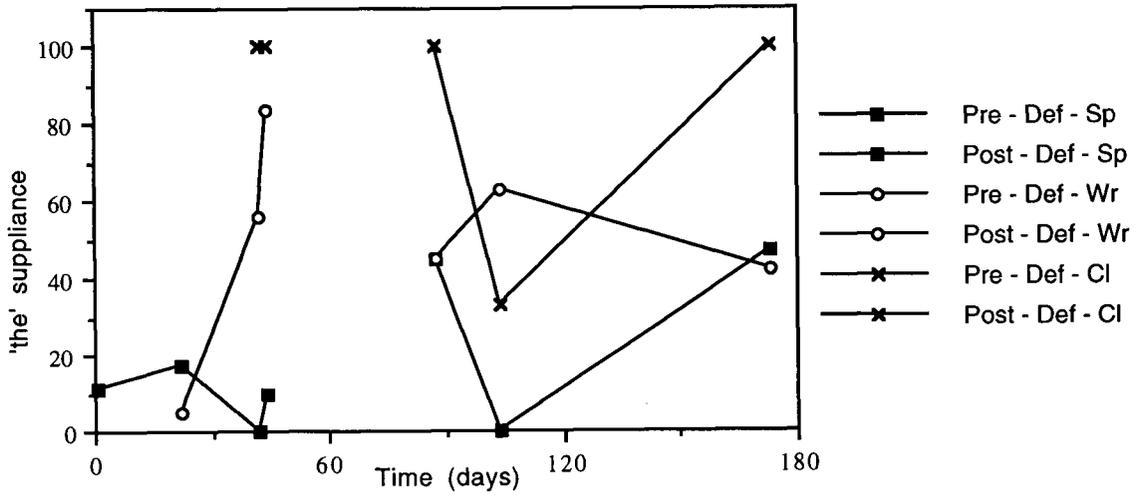
Appendix B5: Learner Con1, Line Graphs of Developmental Course, For All Dependent Variables on All Tasks, With Lines Connecting the Data Points Within Each Phase



Appendix B6: Learner Con2, Line Graphs of Developmental Course, For All Dependent Variables on All Tasks, With Lines Connecting the Data Points Within Each Phase



Appendix B7: Learner Con3, Line Graphs of Developmental Course, For All Dependent Variables on All Tasks, With Lines Connecting the Data Points Within Each Phase



Appendix B8: Learner Con4, Line Graphs of Developmental Course, For All Dependent Variables on All Tasks, With Lines Connecting the Data Points Within Each Phase

Appendix C

The 11 Aesop's Fables Used in the Study

Fable #1

A crow sat in a tree holding in his beak a piece of meat that he had stolen. A fox which saw him decided to get the meat. It stood under the tree and began to tell the crow what a beautiful big bird he was. He ought to be the king of all the birds, the fox said; and he would undoubtedly have been made king, if only he had a voice as well. The crow was so anxious to prove that he had a voice that he dropped the meat and sang for all he was worth. Up ran the fox, snapped up the meat and said to him: "If you added brains to all your other qualifications, you would make an ideal king."

Moral: We should not trust people who flatter us too much.

Fable #2

A donkey crossing a river with a load of salt lost his footing and slipped into the water, so that the salt dissolved. He was very pleased at finding himself relieved of his burden when he got upon his legs again. So the next time he came to a river with a load on his back, thinking that the same thing would happen if he got into the water, he let himself go under on purpose. But this time he was loaded with sponges, which absorbed so much water that he could not keep his head up and was drowned.

Moral: If you are lazy and try to get out of your work, you may have troubles.

Aesop's Fables - #3

A mouse ran over the body of a sleeping lion. Waking up, the lion seized the mouse and was going to eat it. The mouse begged to be released and promised to repay the lion if the lion would free him. The lion laughed and freed the mouse. Soon afterwards, the lion was captured by some hunters and tied by a rope to a tree. The mouse heard the cries of the lion and quickly ran to the tree. The mouse gnawed on the rope and freed the lion.

Fable 4

A horse and a donkey were on a journey with their owner. The small donkey was carrying the same number of bags as the large horse. The donkey became very tired and weak, so the donkey asked the horse to carry some bags in order to save his life. The horse would not help and later the donkey became exhausted, fell down and died. The owner then put all of the donkey's bags and the donkey's body on the horse's back. Although the horse was big and strong, eventually he too became exhausted, fell down and died.

Moral: Those who are strong should help those who are weak so that both will be able to live.

Fable #5

A fox stole a chicken from a farmer and ate the chicken. The farmer became very angry and so when the farmer caught the fox, he decided he would make it pay dearly. The farmer tied a cloth soaked in oil to the fox's tail and set it on fire. In pain, the fox ran into the farmer's corn-fields, which were ready for harvesting. The farmer chased the fox, but the corn caught fire and burned so that the farmer lost all of his harvest.

Moral: We should avoid uncontrolled rage and inhumanity to others, since these may often cause harm to those who hurt others.

Fable #6

A man who was making a long journey vowed to give to charity one half of anything he found on the way. One day he found a small bag. When the man searched the bag, he found only almonds and dates. As he was eating them, the man decided that his vow would be honored since he would give to charity the insides of the dates, the pits, and the outsides of the almonds, the shells. As he bit into one of the dates, he hurt his tooth. Then, the man choked on one of the almonds, suffocated and died.

Moral: We should not be greedy and that we should always honour our vows.

Fable #7

A thirsty deer came to a stream, and after drinking noticed his reflection in the water. He saw his great antlers and was proud. Then he saw his thin weak-looking legs and was unhappy with them. Just then a lion appeared and chased the deer. Because the deer's legs were so fast, the deer easily ran away from the lion. However, the deer soon ran into a forest where his antlers became tangled in the branches of a tree. The lion caught the deer and killed it.

Moral: Do not be vain and so concerned about appearances.

Fable #8

An old man bought a dog and a cat. He trained the dog to hunt and kept the cat as a house-pet. Whenever the dog caught a rabbit or some other game, the cat was also given some of the meat. One day the old man died and the animals were left alone. The dog was able to find food for himself. However, the cat stayed in the man's house and could not catch any food. Eventually the cat became sick and died.

Moral: Parents should not let their children be lazy and idle. Children need to learn to look after themselves.

Fable #9

A greedy old man, a miser, sold all of his possessions, including his warm clothes, and bought a large diamond. The man buried the diamond under a boulder on the top of a hill. Every day, the man walked to the hill and sat on the boulder. A farmer who saw the man each day guessed the man's secret, dug up the diamond and took it away. The old man continued to sit on the stone each day until one year later when he became very cold during a storm, got sick and died.

Moral: We should enjoy the things that we own, otherwise we may as well not have them.

Fable #10

A boar (a wild pig) and a horse lived together in a large meadow. The boar ate a large amount of grass and the horse thought that the boar ate too much grass. The horse became angry and asked a hunter to kill the boar. The hunter said that he could kill the boar only if he was riding the horse. The horse agreed and so the hunter put a bridle on the horse and mounted on its back. After killing the boar, the hunter kept the horse. He tied the horse in a stable and only gave it old straw to eat.

Moral: Anger and greed makes people blind and may cause them to do things they will regret.

Fable #11

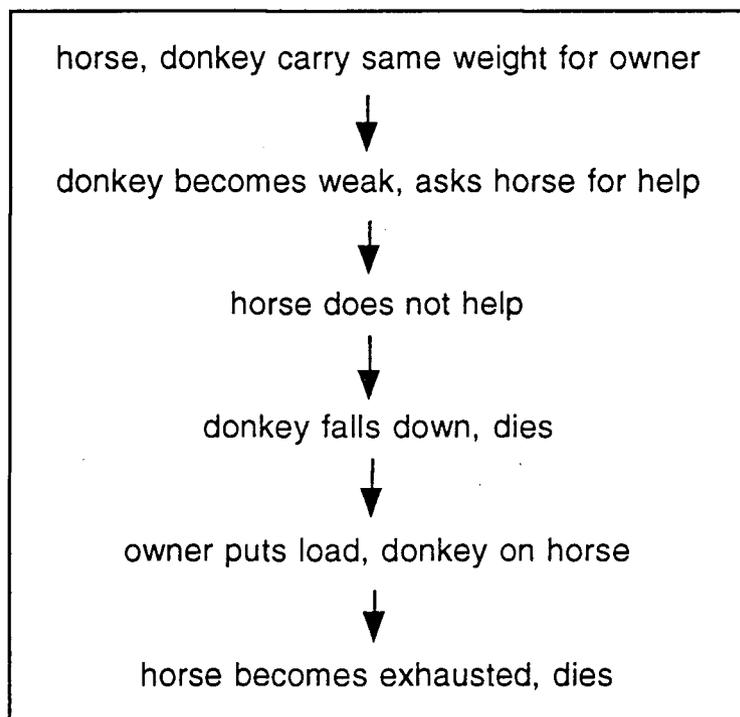
An old farmer knew that he would die soon and wanted to teach his sons to be good farmers. So the farmer said to his sons "I will soon leave this world. My treasure is hidden in our vineyard." The sons thought there was a treasure buried somewhere in the vineyard, and after the father died, they carefully dug every inch of soil around the vines in the vineyard. The sons did not find any treasure. However, because the ground had been so well dug, the vines yielded a bumper crop and the sons became wealthy.

Moral: Our greatest treasure is the fruit of our labours, the things we earn with hard work.

Appendix D

Time Sequence Graphic, Example for One Fable

Fable #3



Appendix E

Composition Worksheet, Example from One Learner (Tr3)

Name _____

In one paragraph, summarize the fable and its moral. Please write by hand on this sheet. Try to take less than 15 minutes. You do not need to look up the spelling of words in a dictionary.

A horse and a donkey were on the trip with their owner. Both horse and donkey carried same amount of bags. The donkey became tired and felt weak, so it asked horse for carrying its bags. However horse did not help donkey. The donkey fell down and died. The owner put all bags which were carried by donkey on the horse and also donkey's dead body. Horse became exhausted and also died.

Whoever the strong should help the weak otherwise in the end both the strong and the weak will die.

Appendix F**Cloze Worksheet, Example from One Fable**

Name _____

Aesop's Fables - #3

In the following fable, please fill in each blank (_____) with an appropriate word, part of a word, or Ø if nothing is required.

_____ horse and _____ donkey were on _____ journey with their owner. The small donkey was carrying the same number of _____ bags as the large horse. _____ donkey became very tired and weak, so the donkey ask_____ the horse to carry some bag_____ in order to save_____ his life. The horse did not help and later the donkey became _____, fell down and died. _____ owner then put all of the donkey's bag_____ and the donkey's body on the horse_____ back. Although _____ horse was big and strong, eventual_____ he too became exhausted, fell down and died.

Appendix G

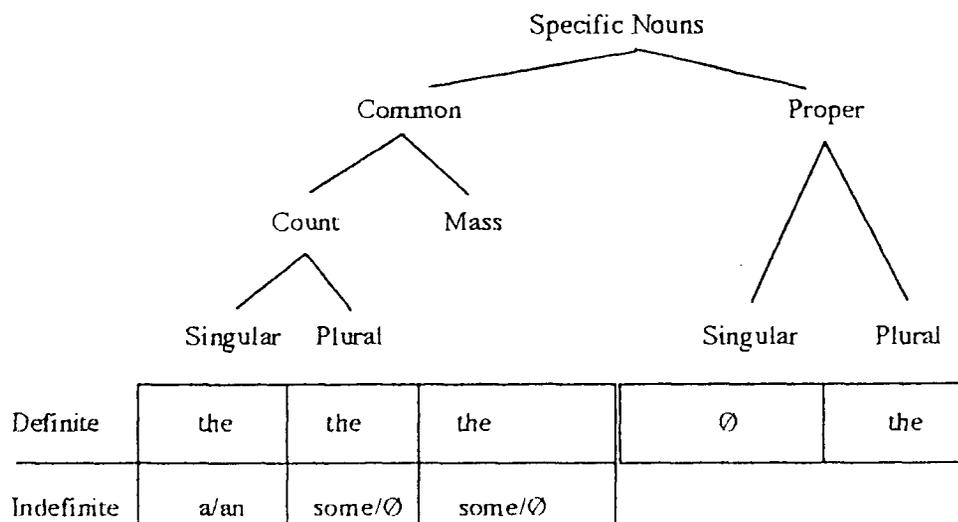
Article Instruction, Grammar Presentation Sheet

English Article Use

Article use is very complex because there are only three choices (the, a or \emptyset) and each is used in several different situations. The article that you will use depends on the context and meaning, not on the specific noun. There are four main categories to know: definite vs indefinite, generic vs specific, common vs proper and count vs mass. Using these categories, there are four basic rules that should help you in most cases.

1. If it is a proper noun and singular, use \emptyset . If it is a proper noun and plural use the.
 e.g. \emptyset Canada is the ^{2nd} largest country in the world.
 \emptyset Ethnocentrism is the belief that one's own patterns of behavior are always natural, good, beautiful or important.
The United States has a new leader.
The Clintons live in the White House.
2. If both you and your reader know the specific thing or person that the noun refers to (i.e. it is definite), use the. This happens most often when the thing has already been mentioned or if you provide information that defines the specific thing.
 e.g. A major debate occurred in Parliament last week. The debate last for two days.
The products that Japan produces are exported to other Asian countries.
3. If you and your reader do NOT both know the specific thing or person that the noun refers to (i.e. the noun is indefinite), use a or \emptyset . This happens most often when the noun is mentioned for the first time.
 e.g. The people that were interviewed provided \emptyset interesting answers.
 A book is on the table. The book is open to page 212.
 There is \emptyset water on the floor. The water is from melted snow. (non-count noun)
4. If you are referring to a class of objects or to an unspecified number of things (i.e. the noun is generic), you are often best to use \emptyset and the plural. You may use the when you are referring to groups of people. If the noun is a mass noun, including abstract nouns, you should use \emptyset and do NOT make the noun plural.
 e.g. I love to eat \emptyset cookies. \emptyset Immigrants to Canada often live in the big cities.
The Mormons mainly live in Utah.
 An economy without \emptyset change is an economy without \emptyset growth.

The following chart explains these factors for SPECIFIC nouns:



1.0 Deciding if a specific noun is definite (i. e. to use the)

Definite: means both the speaker and the hearer know the specific referent of the noun

Some examples of when the referent is definite:

1. When you refer back to a previously mentioned noun, e.g. I saw a dog. THE dog is. .
2. When you refer towards specific information, e.g. THE house *with a view*, THE last sentence *on the page*, I found THE book *that was missing*.
3. When you refer physically to a specific object, e.g. THE floor, THE chair (pointing)
4. When something is unique, e.g. THE moon.

We will practice (1) today.

2.0 Generic Uses - describing a group or class, not a specific individual or entity; not a real or actual noun, but the idea or concept of a noun

1. THE Canadian is a good citizen. - With countable nouns; formal usage; more abstract; for classes of humans, animals, plants, organs, complex devices.
2. THE Canadians are good citizens. - With countable nouns; formal usage; abstract; for classes of humans that are religious, political, national, social, or professional in nature.
3. ___ Canadians are good citizens. - With countable nouns; less formal usage; less abstract; same classes as 1 plus simple inanimate objects.
4. A Canadian is a good citizen. - With countable nouns; colloquial; concrete
5. ___ Rice is a staple food. - With mass nouns.

Appendix H

Scoring Chart, Example from One Learner (Tr4), Spoken Fable #4

Subject (Name) _____

Spoken Composition _____

Date 12/02 Fable number ~~123~~ donkey/horse (4)

Coded by _____ Recoded by _____

CONTEXT	ELEMENT							Total
	the + N	Ø + N	a + N	TN that this	TN one	pronoun Ø	other	
Current/definite-SUB		+++ 1						0/6
Recent/ definite-SUB		1						0/1
Old/ definite - SUB (list N of clauses)	1							0/1
Total/ definite - SUB	1	7						4/8
Current/definite-OBJ	1 11	1						2/3
Recent/ definite -OBJ	1 1							2/2
Old/ definite - OBJ (list N of clauses)								
Total /definite - OBJ	4	1						4/5
<u>cut</u> Total /definite	5	10						5/15
New/ indefinite-SUB		11						
New/ indefinite-OBJ								
Total/new								
Ø (plural, mass)								
TOTAL								

skip ignore possessives

Appendix I

Results for All Learners, All Dependent Variables, Means

Appendix I1 Scores For All Learners and Group Mean Scores for *the* Suppliance for Referential Definites on All Three Tasks, With PreI Phase, Post I Phase, and Gain Scores

		Learners									
		Tr1	Tr2	Tr3	Tr4	TrX	Con1	Con2	Con3	Con4	ConX
Spoken	pre-	40	75	46	50	52.8	21	27	30	10	22
	post-	60	96	73	90	79.8	29	25	84	31	42.3
	gain	20	21	27	40	27	8	-2	54	21	20.3
Written	pre-	57	68	59	73	64.3	58	86	39	46	57.3
	post-	71	98	94	95	89.5	33	79	79	51	60.5
	gain	14	30	35	22	25.3	-25	-7	40	5	3.3
Cloze	pre-	100	100	93	100	98.3	63	100	100	100	90.8
	post-	72	100	93	100	91.3	56	100	100	78	83.5
	gain	-28	0	0	0	-7	-7	0	0	-22	-7.3

Appendix I2 Scores For All Learners and Group Mean Scores for *the* Suppliance in Subject (S) and Non-subject (NS) Positions on the Spoken and Written Tasks, With PreI Phase, Post I Phase, and Gain Scores

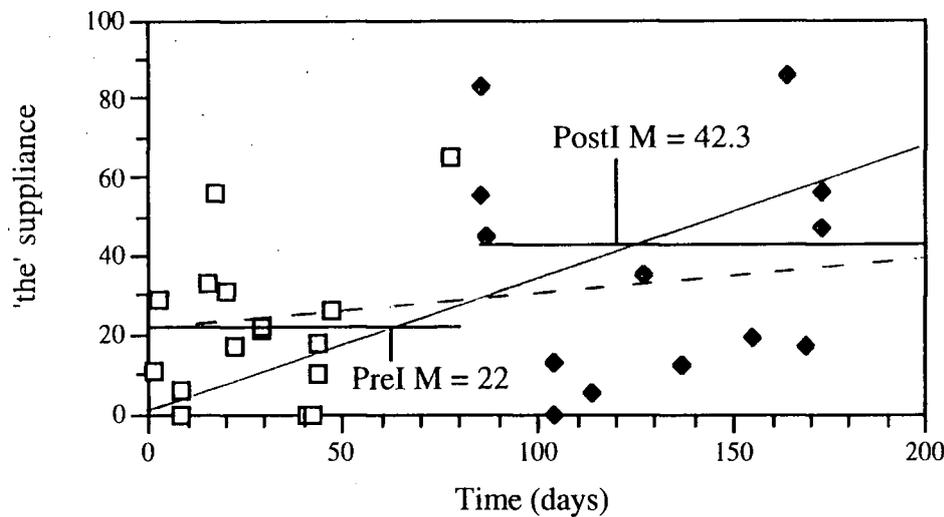
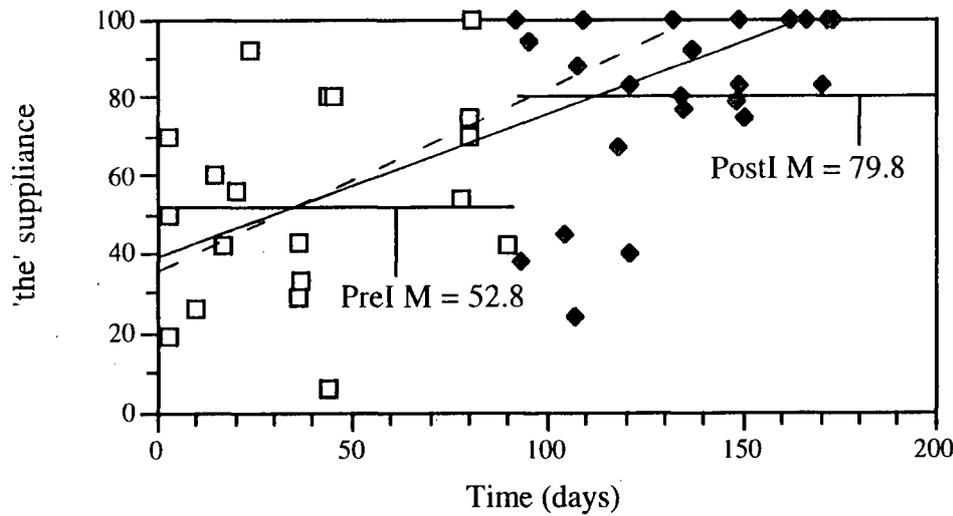
		Learners														
		Tr1			Tr2			Tr3			Tr4			TrX		
		S	NS	Diff	S	NS	Diff	S	NS	Diff	S	NS	Diff	S	NS	Diff
Spoken	Pre-	4	86	-82	59	84	-25	24	57	-33	33	83	-50	30	77.5	-47.5
	Post	38	78	-40	97	96	1	70	77	-7	92	88	4	74.3	84.8	-10.5
	Gain	34	-8	42	38	12	26	46	20	26	59	5	54	44.3	7.3	37
Written	Pre-	39	70	-31	87	69	18	55	65	-10	82	73	9	65.8	69.3	-3.5
	Post	76	79	-3	100	100	0	100	88	12	97	96	1	93.3	90.8	2.5
	Gain	37	9	28	13	31	-18	45	23	22	15	23	-8	27.5	21.5	6
Spoken		Con1			Con2			Con3			Con4			ConX		
		S	NS	Diff	S	NS	Diff	S	NS	Diff	S	NS	Diff	S	NS	Diff
	Pre-	3	38	-35	14	48	-34	3	64	-61	0	27	-27	5	44.3	-39.3
Post	44	24	20	11	48	-37	78	90	-12	32	31	1	41.3	48.3	-7	
Gain	41	-14	55	-3	0	-3	75	26	49	32	4	28	36.3	4	32.3	
Written	Pre-	75	48	27	100	80	20	29	53	-24	55	39	16	64.8	55	9.8
	Post	70	25	45	100	70	30	89	76	13	74	35	39	83.3	51.5	31.8
	Gain	-5	-23	18	0	-10	10	60	23	37	19	-4	23	18.5	-3.5	22

Appendix I3 Scores For All Learners and Group Mean Scores for *the* Suppliance for Current (C) and Recent (R) Referents on the Spoken and Written Tasks, With PreI Phase, Post I Phase, and Gain Scores

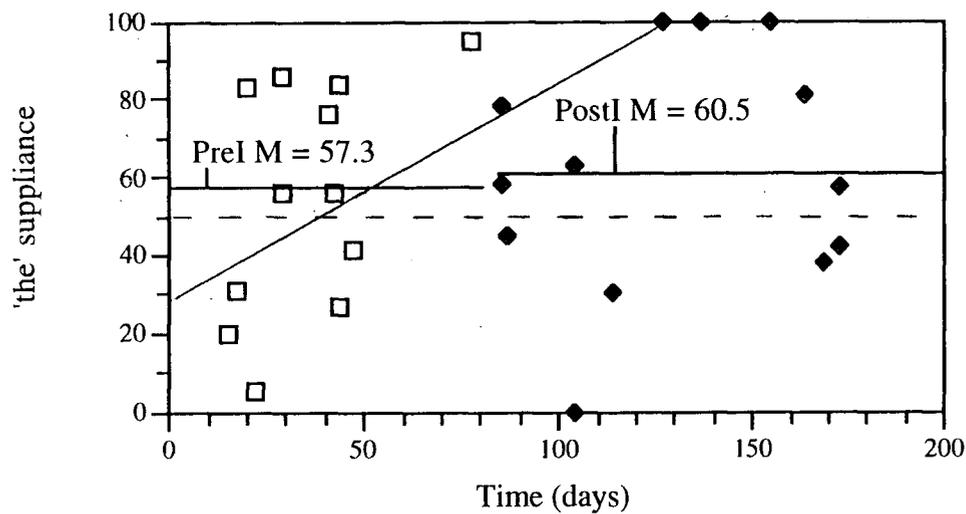
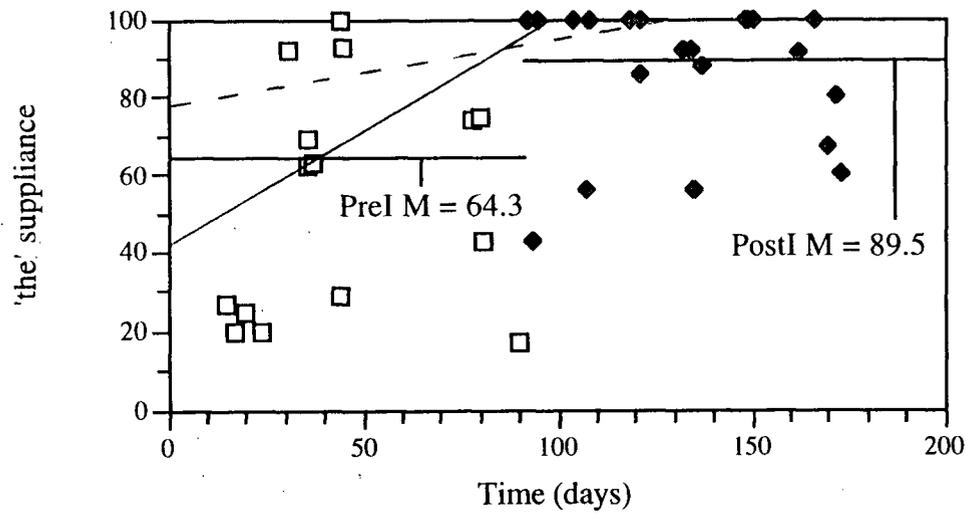
		Learners														
		Tr1			Tr2			Tr3			Tr4			TrX		
		C	R	Diff	C	R	Diff	C	R	Diff	C	R	Diff	C	R	Diff
Spoken	Pre-	32	42	-10	71	75	-4	35	44	-9	44	65	-21	45.5	56.5	-11
	Post	51	67	-16	100	93	7	82	65	17	88	92	-4	80.3	79.3	+1
	Gain	19	25	-6	29	18	11	47	21	26	44	27	17	34.8	22.8	12
Written	Pre-	47	61	-14	75	79	-4	70	54	16	79	80	-1	67.8	68.5	-0.8
	Post	73	85	-12	100	100	0	100	93	7	96	100	-4	92.3	94.5	-2.3
	Gain	26	24	2	25	21	4	30	39	-9	17	20	-3	24.5	26	-1.5
		Con1			Con2			Con3			Con4			ConX		
		C	R	Diff	C	R	Diff	C	R	Diff	C	R	Diff	C	R	Diff
Spoken	Pre-	18	18	0	22	35	-13	21	36	-15	8	7	1	17.3	24	-6.8
	Post	15	55	-40	27	21	6	87	82	5	26	36	-10	38.8	48.5	-9.8
	Gain	-3	37	-40	5	-14	19	66	46	20	18	29	-11	21.5	24.5	-3
Written	Pre-	59	59	0	89	100	-11	29	54	-25	61	41	20	59.5	63.5	-4
	Post	62	25	37	93	64	29	79	92	-13	53	58	-5	71.8	59.8	12
	Gain	3	-34	37	4	-36	40	50	38	12	-8	17	-25	12.3	-3.8	16

Appendix J

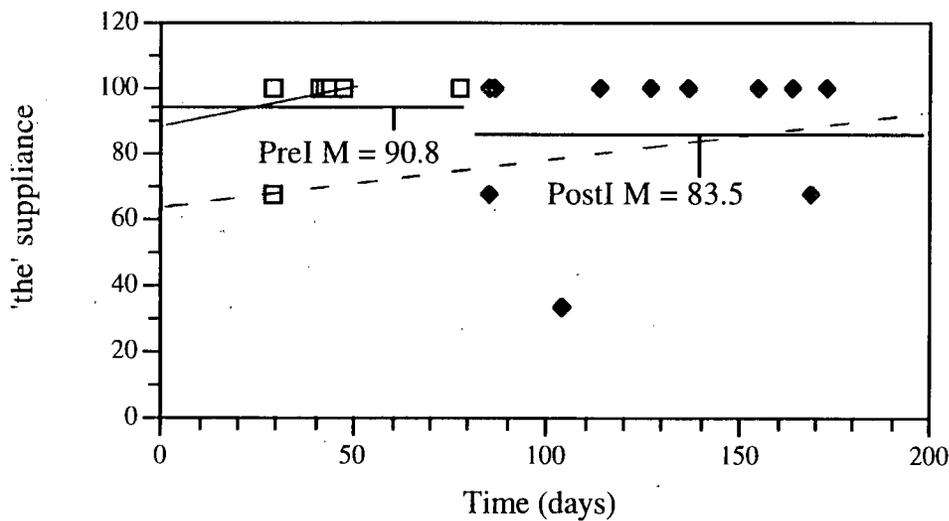
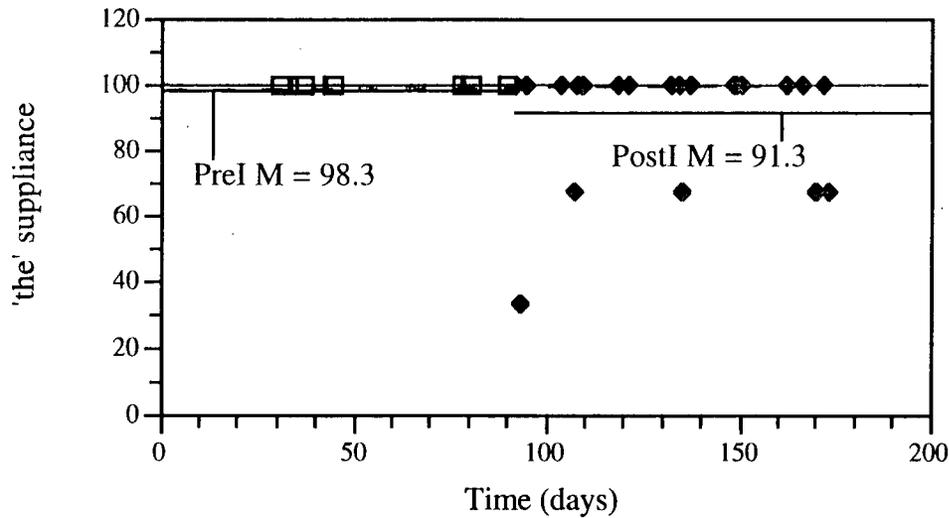
Group Regression Lines for All Dependent Variables



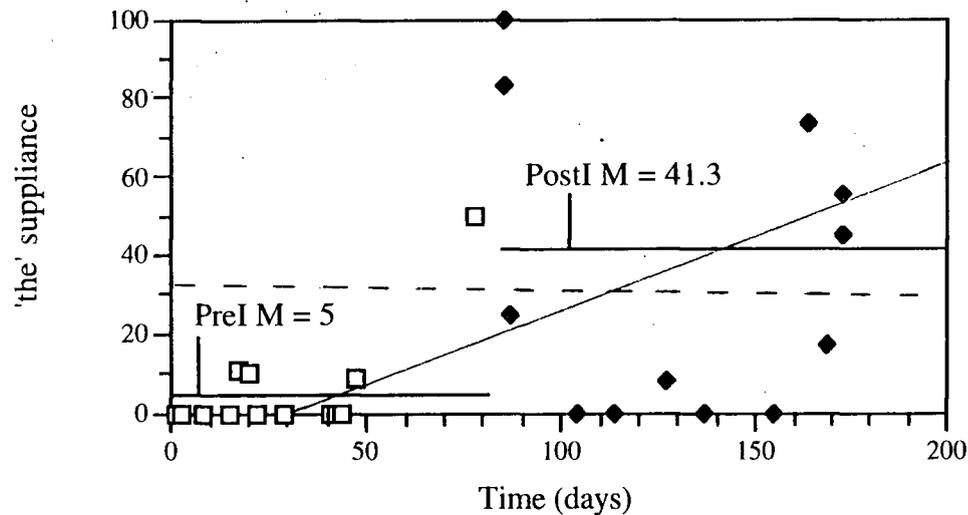
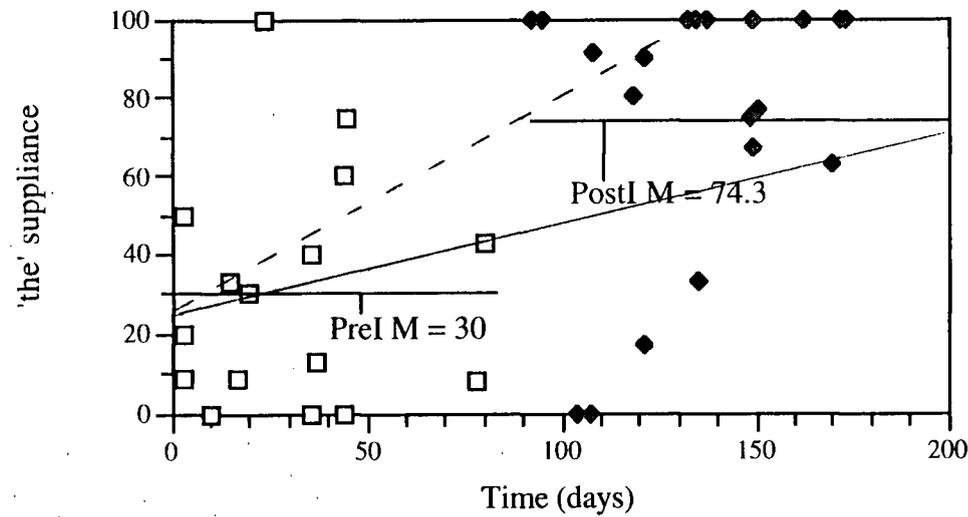
Appendix J1: Percentage of *the* Suppliance for All Referential Definite Noun Phrases on Spoken Retelling Tasks, for the Treatment (T) and Control (C) groups, with a Mean Group Score Line (M) for Each Phase (PreI and PostI), and with Linear Regression Lines for Each Phase Extending Throughout the Entire Study Period.



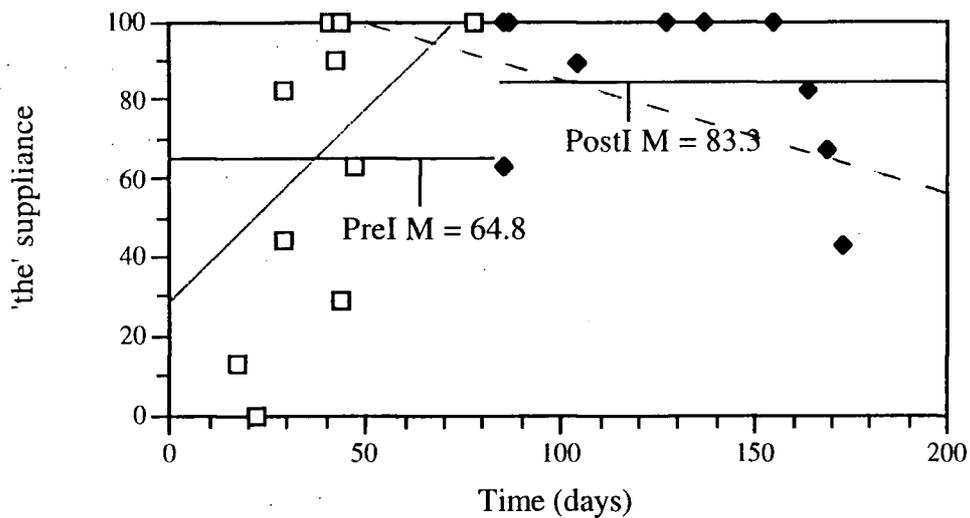
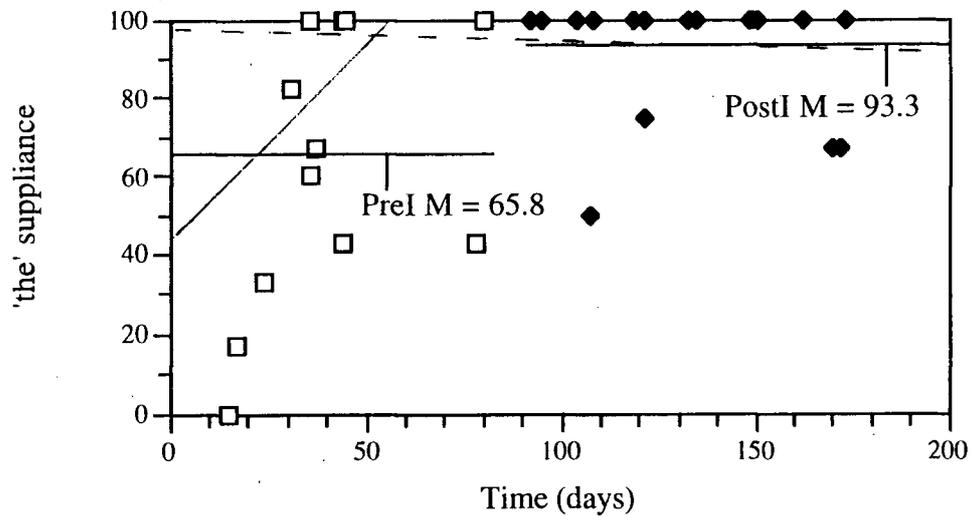
Appendix J2: Percentage of *the* Suppliance for All Referential Definite Noun Phrases on Written Retelling Tasks, for the Treatment (T) and Control (C) groups, with a Mean Group Score Line (M) for Each Phase (PreI and PostI), and with Linear Regression Lines for Each Phase Extending Throughout the Entire Study Period.



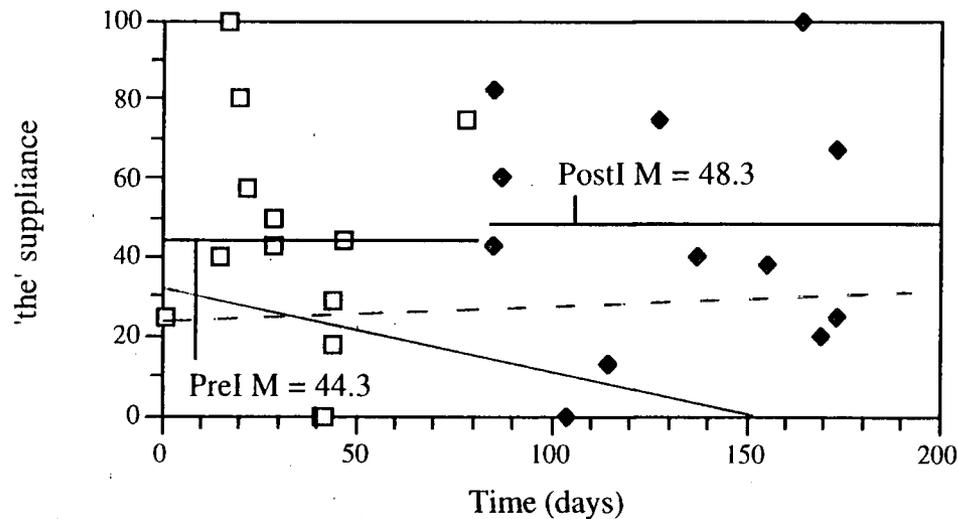
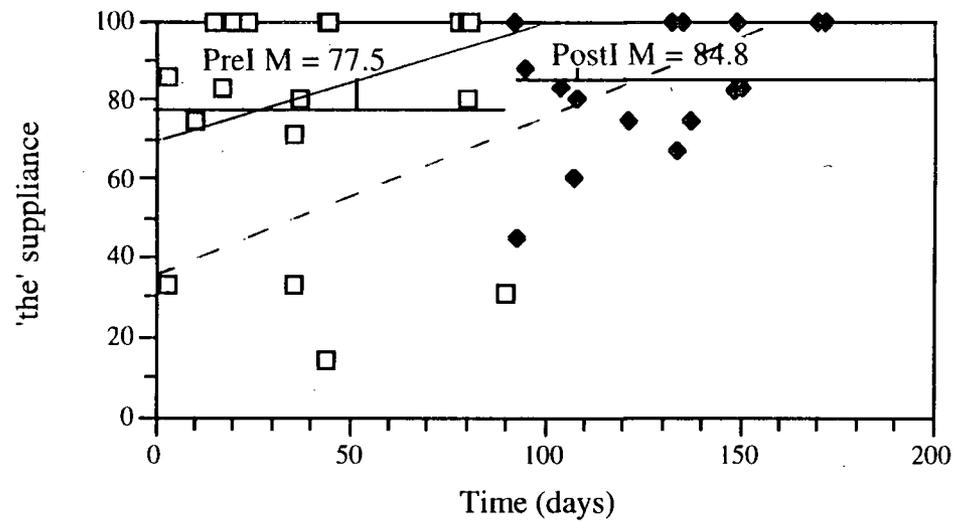
Appendix J3: Percentage of *the* Suppliance for All Referential Definite Noun Phrases on Cloze Retelling Tasks, for the Treatment (T) and Control (C) groups, with a Mean Group Score Line (M) for Each Phase (PreI and PostI), and with Linear Regression Lines for Each Phase Extending Throughout the Entire Study Period.



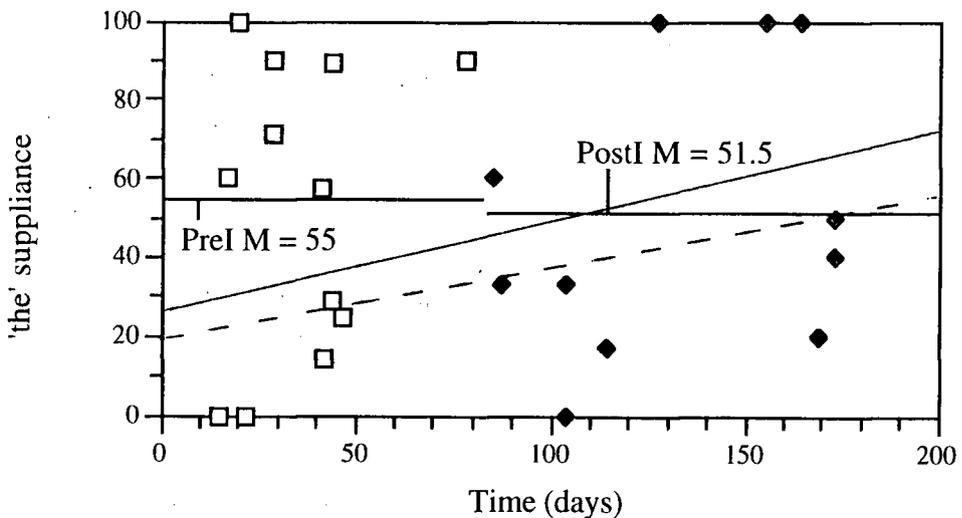
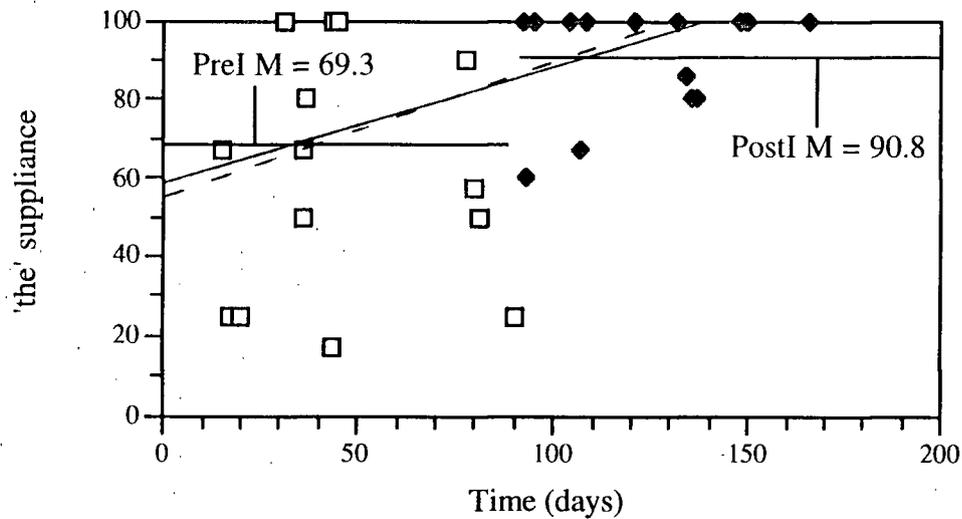
Appendix J4: Percentage of *the* Suppliance in Subject Positions on Spoken Retelling Tasks, for the Treatment (T) and Control (C) groups, with a Mean Group Score Line (M) for Each Phase (PreI and PostI), and with Linear Regression Lines for Each Phase Extending Throughout the Entire Study Period.



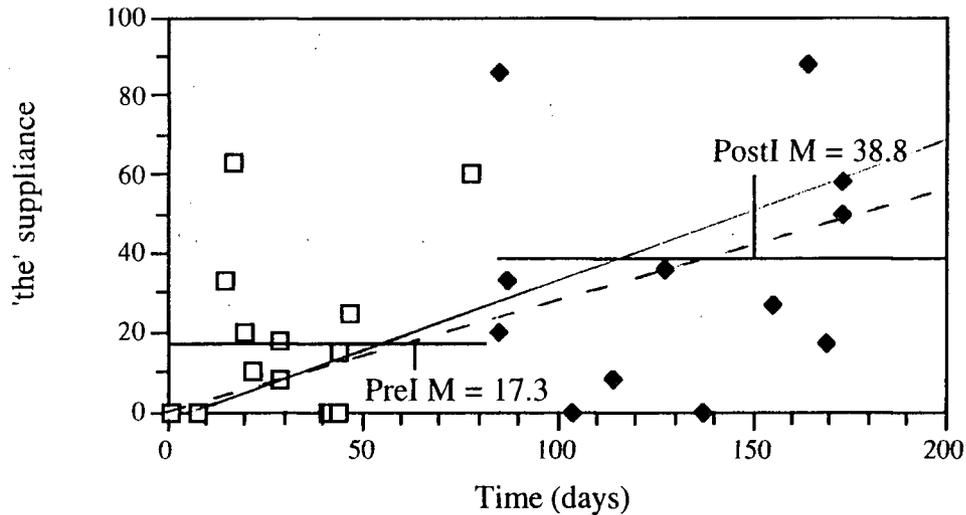
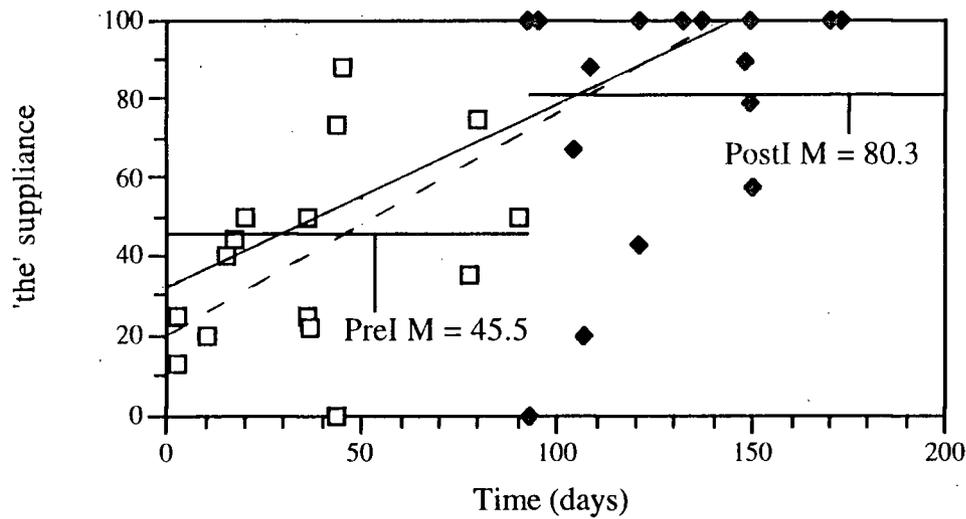
Appendix J5: Percentage of *the* Suppliance in Subject Positions on Written Retelling Tasks, for the Treatment (T) and Control (C) groups, with a Mean Group Score Line (M) for Each Phase (PreI and PostI), and with Linear Regression Lines for Each Phase Extending Throughout the Entire Study Period.



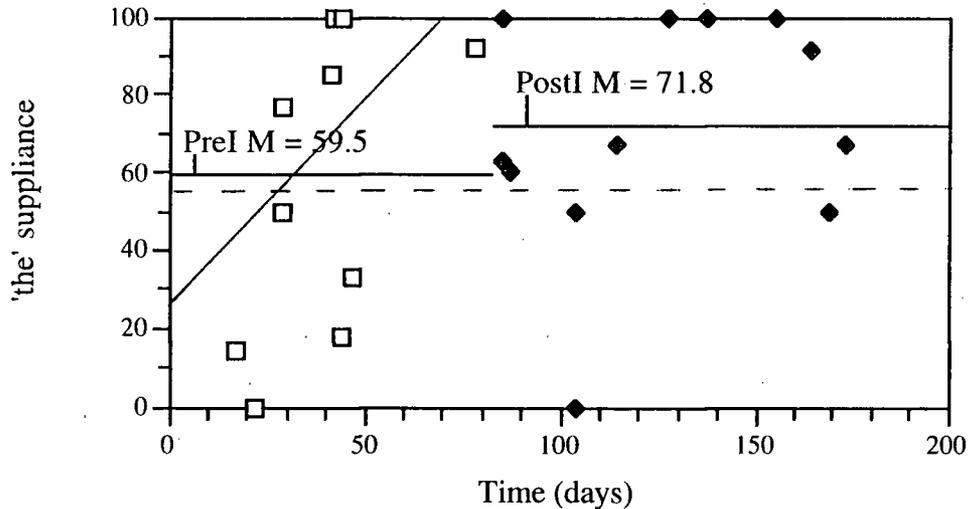
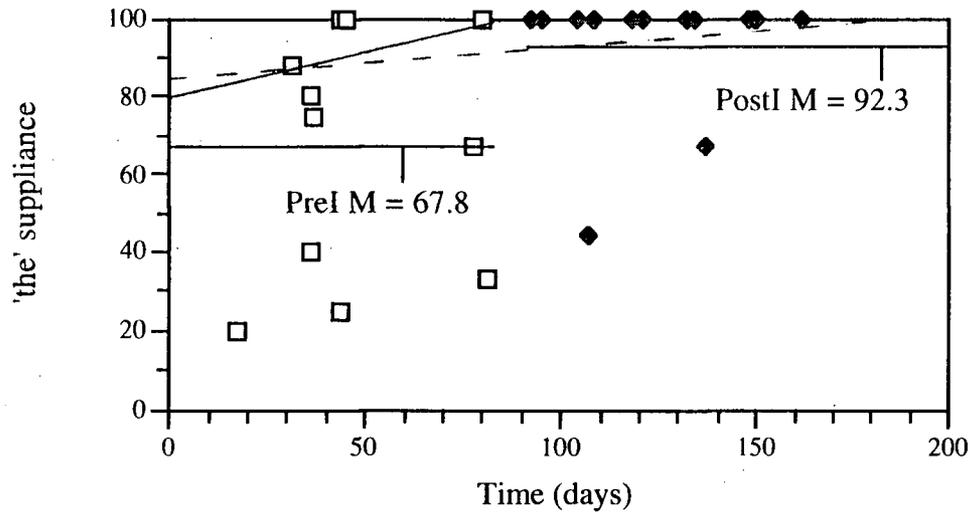
Appendix J6: Percentage of *the* Suppliance in Non-subject Positions on Spoken Retelling Tasks, for the Treatment (T) and Control (C) groups, with a Mean Group Score Line (M) for Each Phase (PreI and PostI), and with Linear Regression Lines for Each Phase Extending Throughout the Entire Study Period.



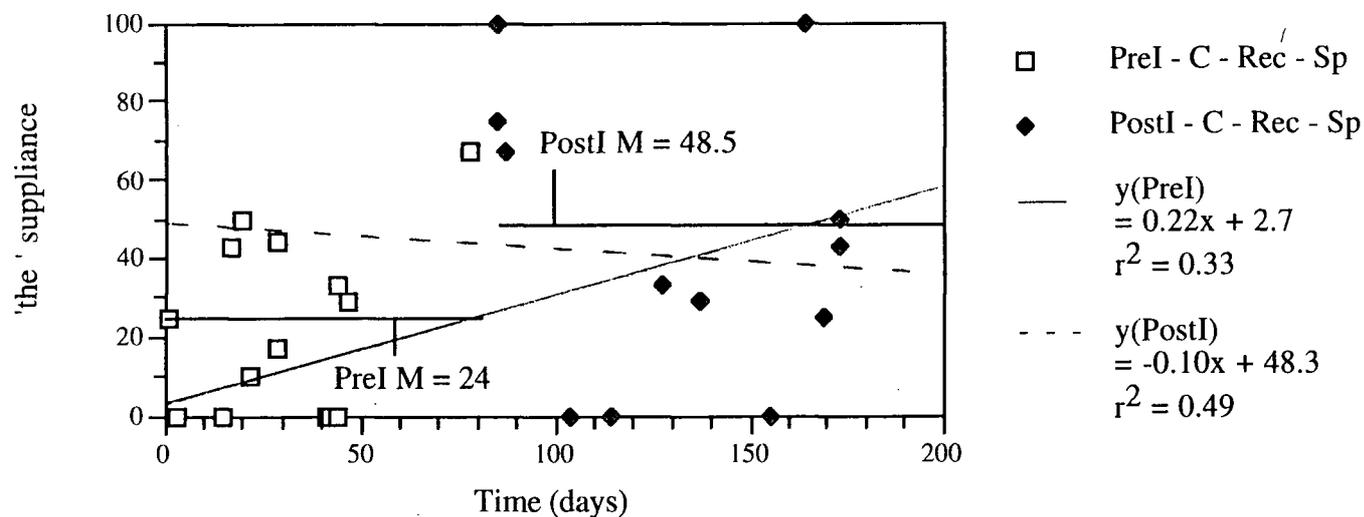
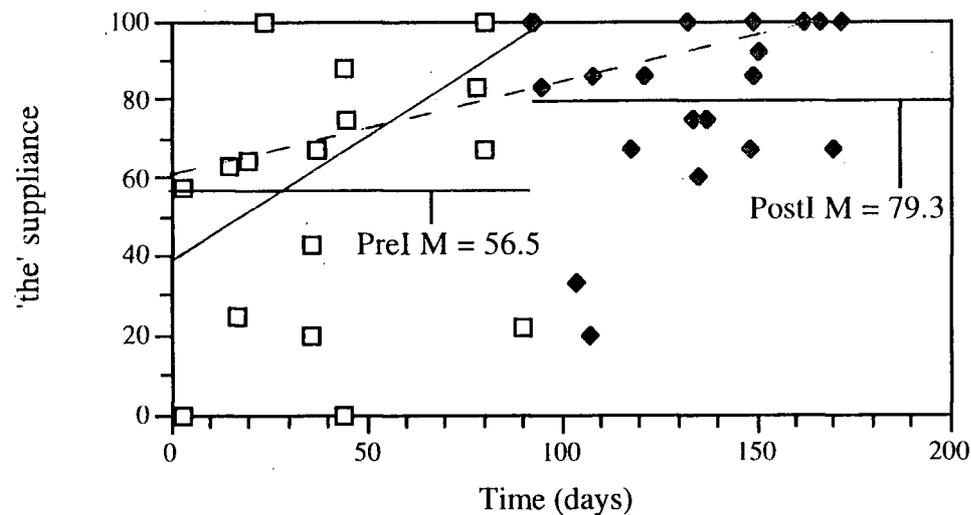
Appendix J7: Percentage of *the* Suppliance in Non-subject Positions on Written Retelling Tasks, for the Treatment (T) and Control (C) groups, with a Mean Group Score Line (M) for Each Phase (PreI and PostI), and with Linear Regression Lines for Each Phase Extending Throughout the Entire Study Period.



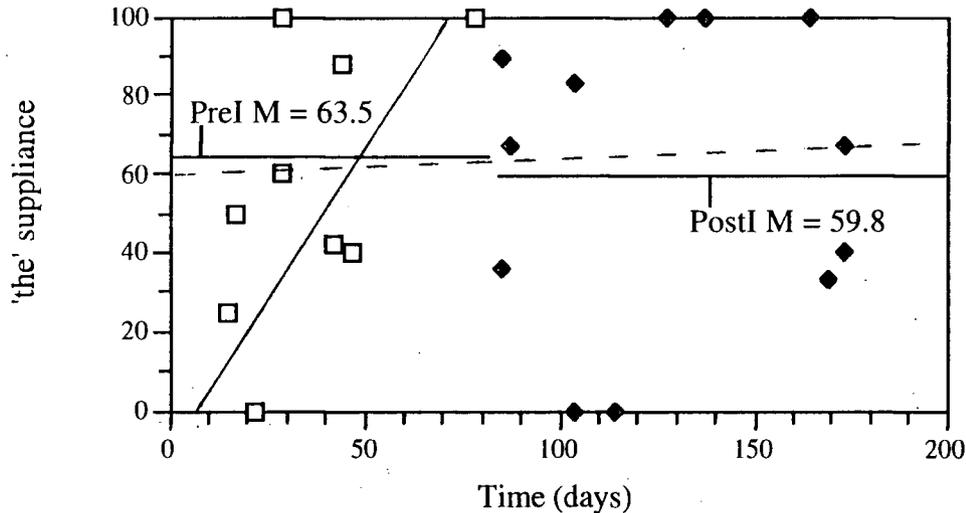
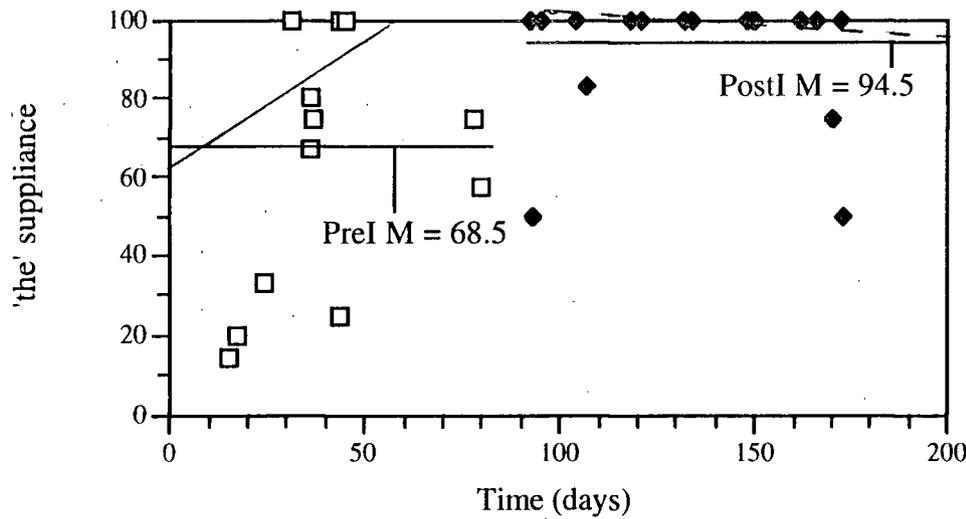
Appendix J8: Percentage of *the* Suppliance for Current Referents on Spoken Retelling Tasks, for the Treatment (T) and Control (C) groups, with a Mean Group Score Line (M) for Each Phase (PreI and PostI), and with Linear Regression Lines for Each Phase Extending Throughout the Entire Study Period.



Appendix J9: Percentage of *the* Suppliance for Current Referents on Written Retelling Tasks, for the Treatment (T) and Control (C) groups, with a Mean Group Score Line (M) for Each Phase (PreI and PostI), and with Linear Regression Lines for Each Phase Extending Throughout the Entire Study Period.



Appendix J10: Percentage of *the* Suppliance for Recent Referents on Spoken Retelling Tasks, for the Treatment (T) and Control (C) groups, with a Mean Group Score Line (M) for Each Phase (PreI and PostI), and with Linear Regression Lines for Each Phase Extending Throughout the Entire Study Period.



Appendix J11: Percentage of *the* Supplience for Recent Referents on Written Retelling Tasks, for the Treatment (T) and Control (C) groups, with a Mean Group Score Line (M) for Each Phase (PreI and PostI), and with Linear Regression Lines for Each Phase Extending Throughout the Entire Study Period.