TIME DISTRIBUTION AND THE ACQUISITION OF ENGLISH AS A FOREIGN LANGUAGE

Tesi doctoral presentada per

Raquel Serrano Serrano

com a requeriment per a l'obtenció del títol de

Doctora en Filologia Anglesa

Programa de Doctorat: *Lingüística Aplicada*

Bienni 2002-2004

Departament de Filologia Anglesa i Alemanya

Universitat de Barcelona

Setembre 2007

Directora: **Dra. Carme Muñoz Lahoz**

"We need to create a mindset in which both teachers and researchers view classrooms as laboratories where theory and practice can interact to make both better practice and better theory."
Schachter (1993: 181)

ACKNOWLEDGMENTS

This Ph.D. dissertation would not have been possible without the help of many people who were involved in it. These few paragraphs are dedicated to all of them.

First and foremost, I would like to thank my supervisor, Dr. Carme Muñoz. I could certainly write a whole book thanking Dr. Muñoz, and this brief paragraph does not do justice to all my gratitude and admiration towards her. It has been a privilege to benefit from Dr. Muñoz's ample knowledge and expertise in the area of SLA, her very insightful comments, her guidance, and her feedback. Not only has Dr. Muñoz been a crucial reference in terms of research, she has also impressed me with her personal qualities, without which, this dissertation would not have been completed in such good spirits. She has been incredibly supportive, fast, efficient, and dedicated, despite all the other issues which have constantly demanded her attention. It has indeed been a pleasure and an honor (both personally and professionally) to work with Dr. Carme Muñoz.

My thanks are extended to Dr. Laura Collins and Dr. Joana White for their seminar at the UB during the academic year 2003-2004. It was thanks to them that my interest in intensive L2 instruction arose, which was the seed for this dissertation. I also thank Dr. Patsy Lightbown for her seminar during the spring semester 2007, and for her very insightful comments about aspects of my research. My appreciation additionally goes to those prominent researchers who have given me punctual feedback and answered my questions promptly, some of them without even knowing me (Dr. Robert DeKeyser, Dr. Robert Greene, Dr. Philip Pavlik, Dr. Thomas Toppino, and Dr. Michael Ullman).

I would also like to thank the Escola d'Idiomes Moderns for allowing me to collect data from their students. I am grateful to the director, Dr. Philip Banks, and especially to the coordinators (Paul Ambrose, Richard Blair and Jon Gregg), for their help and support. I am indebted to the teachers who participated in my study and above all to the students. Without their participation this research would have been impossible.

I am indebted to all the members of the GRAL research team, who have

supported and helped me during these years. Thanks to Dr. Roger Gilabert for sharing his knowledge with respect to crucial areas in this dissertation, especially in terms of measures of oral language production (which were of primary importance in my research), and for his encouragement and readiness to help. I am thankful to Imma Miralpeix for her insightful ideas about vocabulary, and her willingness to assist in any area, including statistics. I thank Dr. Teresa Navés for her ideas regarding measures of written production. Thanks to Dr. Elsa Tragant, Dr. M. Rosa Torras, and Dr. Joan Carles Mora for sharing relevant articles with me and for their support. Thanks to Dr. María Luz Celaya, Laura Sánchez and Anna Marsol for their encouragement. Additionally, I would like to thank the following research assistants, who helped me in the data collection procedure: Mar Suárez, Gisela Grañena, Cristina Aliaga, and also Júlia Barón and Maria Àngels Llanes, who not only assisted me with work, but also with their invaluable friendship.

Thanks to my oldest friend Cristina for helping me in various technical aspects of this dissertation, but most importantly for her unconditional friendship and support. Thanks to my other friends: Alberto (who has also helped me with my work), Jordi, Montse, Virginia, Andrei, and all the others who have encouraged me during this process and offered me company, love and laughs. Thanks to Joe for bringing everything together.

Me gustaría dedicar esta tesis a mi familia, muy especialmente a mis padres, quienes siempre me han dado ánimos y me han apoyado para llegar hasta aquí.

A Petra y Benjamín

ABSTRACT

The purpose of the present dissertation is to investigate the effect of time distribution of practice (or instruction hours) on the acquisition of English as a foreign language. The general objective is to have an overall picture of the learners' language gains in English according to the type of program they attended (intensive or non-intensive), focusing on all the different language areas: listening, grammar, vocabulary, reading, writing and speaking. Additionally, another main objective of analysis is the relationship between the students' initial proficiency level and the effect of time distribution on learners' L2 gains. While there is general agreement that more hours of foreign language instruction (the same as more hours of practice in any skill) leads to more advanced levels of performance, there is not such consensus about how the hours of classroom instruction should be distributed. Traditionally, most institutions providing second language teaching design programs which offer a limited amount of contact hours with the L2 spread over long periods of time (usually years in the case of most L2 classes in primary or secondary school). The present study will analyze data from intermediate and advanced EFL learners in intensive and non-intensive English programs in order to provide some evidence for the optimal distribution of instruction hours in English classes depending on the students' initial proficiency level. Considering previous studies on the effect of time distribution on L2 learning, the first hypothesis predicts that intensive instruction should have a positive effect on the acquisition of English grammar and vocabulary, as well as reading, writing and speaking skills. The second hypothesis states that time distribution should have different effects according to the learners' initial proficiency level. The results of the tests which assessed the students' L2 skills (including a listening activity, a sentence transformation exercise, a cloze activity and a reading) demonstrate that the students in intensive courses seem to make more progress in certain L2 areas (such as listening or grammar), but not in others (reading). However, such advantage was only observed for the students in the intermediate course. Moreover, the analysis of the learners' writing and speaking skills also suggests a certain advantage for the students in intensive courses in some measures of written and oral production, but not in all of them. Similarly, such advantage is only obvious for intermediate learners. These results are interpreted in light of Anderson's ACT model of skill acquisition (Anderson, 1993), which considers that, before a skill is completely acquired the knowledge required to perform such skill moves through several stages, including a declarative stage, a procedural stage and an automatization stage. It will be argued that intensive instruction seems to be more

beneficial when learners are proceduralizing their L2 knowledge (as is the case of the intermediate learners), since the type of input received in intensive classes promotes the acquisition of such knowledge. Advanced learners, on the other hand, need to either acquire new nuances about rules which they have already proceduralized, or eradicate non-target-like productions which have been automatized. It appears that intensive instruction (or at least a short intensive course) does not necessarily help in the acquisition of the knowledge which advanced learners need to acquire. This dissertation concludes by suggesting that the distribution of instruction hours is an important factor in the acquisition of English as a foreign language and should be considered when designing EFL programs.

RESUMEN

El propósito de esta tesis doctoral es investigar el efecto de la distribución de las horas de aprendizaje en la adquisición del inglés como lengua extranjera. El principal objetivo es ofrecer una visión general del progreso que los aprendices realizan en inglés dependiendo del tipo de programa que siguen (intensivo o no intensivo) centrándose en las diferentes áreas: comprensión oral, gramática, vocabulario, lectura, escritura y producción oral. Además, otro objetivo fundamental de este estudio es analizar la relación entre el efecto de la distribución de las horas de aprendizaje y el nivel inicial de inglés de los estudiantes. Mientras que existe un cierto acuerdo entre expertos que cuantas más horas dediquen los estudiantes al aprendizaje de una lengua extranjera más altos niveles de competencia podrán alcanzar, no hay muchos estudios empíricos que se centren en analizar la forma óptima de distribuir las horas de aprendizaje. Tradicionalmente, la mayoría de las escuelas ofrecen un contacto con la lengua extranjera poco intensivo y distribuido en largos períodos de tiempo (años, en el caso de los estudiantes de escuela primaria y secundaria). Esta tesis pretende analizar el progreso de los estudiantes de distinto nivel de inglés (intermedio y avanzado) en distintos tipos de programas (intensivo y no intensivo) para ofrecer evidencia sobre la forma ideal de distribuir las horas de inglés teniendo en cuenta el nivel de los estudiantes. Considerando la investigación realizada en este campo, la primera hipótesis de este estudio afirma que la intensidad tendría que ser un factor positivo en el aprendizaje del inglés. La segunda hipótesis plantea que el efecto de la intensidad tendría que ser diferente para estudiantes con un nivel intermedio y con un nivel avanzado. Los resultados de los tests utilizados para evaluar la

competencia de los estudiantes en inglés (que incluyen una actividad de comprensión oral, dos de gramática y vocabulario y otra de comprensión escrita) parecen sugerir que los estudiantes en cursos intensivos progresan más que los de cursos no intensivos en ciertos aspectos de la lengua (comprensión oral, gramática y vocabulario), pero no en otros (comprensión escrita). Sin embargo, esta ventaja sólo se observa en los grupos de nivel intermedio. Además, los análisis de la producción escrita y oral de los aprendices también sugieren un progreso mayor en el caso de los alumnos en cursos intensivos en algunas medidas utilizadas en este estudio para examinar la producción escrita y oral. De manera parecida, esta ventaja sólo es evidente para los alumnos de nivel intermedio. Estos resultados se interpretarán utilizando el modelo de aprendizaje de Anderson (Anderson, 1993) ACT, según el cual, para adquirir cualquier habilidad (sea hablar una lengua extranjera o conducir un coche) el conocimiento requerido para tal habilidad pasa por las etapas declarativa y procedural antes de ser completamente automatizado. Este estudio sugiere que el aprendizaje intensivo del inglés es más efectivo cuando los estudiantes están proceduralizando su conocimento de la L2 (como es el caso de los estudiantes de nivel intermedio), ya que el *input* que reciben estos estudiantes en cursos intensivos promueve la adquisición del conocimiento procedural. Por otro lado, los estudiantes de nivel avanzado ya han proceduralizado la mayor parte de las reglas del inglés, y en este nivel necesitan aprender nuevas aplicaciones de las reglas que ya conocen, o erradicar formas que distan de la producción de los nativos. Analizando los resultados de este estudio, parece que la intensidad en el aprendizaje no ayuda mucho en esta etapa de adquisición del inglés (o al menos un curso intensivo que ofrece pocas horas). Esta tesis concluye sugiriendo que la distribución de las horas de aprendizaje es un factor importante a tener en cuenta a la hora de diseñar programas de inglés como lengua extranjera y tendría que ser considerado para optimizar el aprendizaje de los estudiantes, dependiendo de su nivel inicial.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

The importance of the time factor in second/foreign language instruction has been emphasized by a number of authors, such as Carroll (1963; 1967; 1989) or Stern (1985). The general claim has been that, all things being equal[1], the more time a student devotes to the learning of a second language, the higher his/her level of proficiency will be (Carroll, 1967; Stern, 1985). Similarly, the longer musicians, sportspeople, painters, etc., spend training and practicing specific skills, the more advanced their performance will become (Ericsson & Charness, 1994; Ericsson, Krampe, & Tesch-Römer, 1993). Essentially, opportunities to practice a particular skill enhance performance, whether such practice involves learning one's native language, a second language, playing the piano, or running a marathon. Moreover, practice usually involves long periods of time before a skill is fully mastered.

Newell & Rosenbloom (1981) explained the maxim which states that "practice makes perfect" with a mathematical power function, according to which, time devoted to practicing a skill is a determining factor in automatization, as reflected by a more accurate and faster performance. In similar lines, in Anderson's ACT theory of human cognition (Anderson, 1993; Anderson et al., 2004), increased practice is what explains the transition from declarative to procedural knowledge,

[1] The "all things being equal" should be highlighted, since research has shown that age is a crucial aspect which affects the amount of time necessary for students to reach a certain proficiency level (Muñoz, 2006; 2007; Swain, 1981; Turnbull et al., 1998). Likewise, time available for learning does not need to correspond to the time which is actually used, as learners may not employ such available time *actively* to learn (Baddeley, 1990).

as well as from procedural knowledge to fully automatized performance.

With respect to language acquisition, both first and second, N. Ellis (2001: 36) has claimed the following:

> Various theories of language acquisition, including connectionist approaches (...), functional linguistics (...), emergentist approaches (...), and cognitive linguistics (...), believe that as the study of language turns out to consider ontogenetic acquisition processes, it favours a conclusion whereby the complexity of the final result stems from simple learning processes applied, *over extended periods of practice in the learner's lifespan* (my italics) to the rich and complex problem space of language evidence.

Although increased time devoted to practicing a skill is commonly claimed to lead to improved performance by cognitive psychologists (Anderson, 1993; Logan, 1988; Newell & Rosenbloom, 1981), as well as by pedagogical experts (Carroll, 1963; 1981), and also by second language acquisition (SLA) researchers (DeKeyser, 2007a; Stern, 1985), the issue of how such time should be distributed is subject to debate. There is a certain agreement that repetition helps learning and "progressively frees the mind from attention to details (…) and reduces the extent to which consciousness must concern itself with the process" (Huey, 1968 as reported in Johnson, 1996: 138). Nevertheless, repetitions need to have certain characteristics in order for them to be effective, such as meaningfulness (in the case of language learning, form-meaning mappings should be reinforced), since rote repetition has been shown to have little effect on learning. Furthermore, it appears that the distribution of the repetitions is an important factor which affects learning.

In the cognitive psychology literature, learning has been shown to improve when the repetitions of the items to be learned appear in distributed sequences, as opposed to massed or concentrated presentations. This phenomenon is known as

the *spacing effect*, according to which including intervals between learning episodes (distributed practice) is more effective for subsequent learning and retention than 'massed' training episodes, in which learning takes place in a concentrated period of time[2]. The spacing effect has been found in a variety of contexts, such as advertisement, mathematical or text processing skills (Appelton, Bjork, & Wickens, 2005; Gay, 1974). As for language, most experiments have analyzed learning of lexical items, whether they include words from the participants' native language, foreign language vocabulary or non-words (Bahrick, 1979; Bahrick & Phelps, 1987; Pavlik & Anderson, 2005; Russo & Mammarella, 2002).

On the other hand, the SLA studies which have investigated the effect of time distribution of instruction hours on students' learning have found that the students tend to make more language gains in intensive foreign language classes as opposed to those classes which offer regular, *drip-feed* instruction, in the case school learners (Collins et al., 1999; Lightbown & Spada, 1994; Netten & Germain, 2004a; White & Turner, 2005), but also in the case of adults (McKee, 1983; Serrano & Muñoz, 2007). Such advantages in favor of students in intensive second language (L2) contexts have been demonstrated for all language skills, although more advantages have been reported in the case of audio-oral skills.

The purpose of the present research study is to investigate the effect of time distribution of practice (or instruction hours) on the acquisition of English as a foreign language. The objective is to have an overall picture of the learners'

[2] Concentration of practice can refer to a massive amount of practice in a short period of time, short time intervals during learning sessions, or short intervals between the presentations of the target items.

language gains in English, focusing on all the different language areas: listening, grammar, vocabulary, reading, writing and speaking. Additionally, another main objective of analysis is the relationship between the students' initial proficiency level and the effect of time distribution on learners' L2 gains.

In Chapter 2, the effects of practice (in terms of time dedicated to activities related to a specific skill, or time-on-task) will be analyzed from the point of view of cognitive psychology (section 2.1). In general, research in this field demonstrates that the automatization of performance in a certain skill requires extensive practice. Next, in section 2.2, the effect of practice in SLA will be examined. The acquisition of an L2 will be assumed to involve the same mechanisms as those responsible for the acquisition of other complex cognitive skills; consequently, practice will also be considered a requirement for automatization of L2 skills. The ideas reported in this chapter will be summarized in section 2.3.

Chapter 3 will explore the role of time distribution in skill acquisition in general and in SLA in particular. Studies investigating the spacing effect will be presented, as well as the possible explanations which have been proposed in the literature in order to account for the occurrence of such phenomenon (section 3.1). Section 3.2 will introduce research on the distribution of instruction hours and the learning of different subjects in high school and college. Afterwards, a literature review on the effect of time distribution on the acquisition of foreign languages will be provided (section 3.3). Section 3.4 will offer a summary of the issues discussed in this chapter.

The research questions which guide this study will be presented in Chapter

4. The first question is concerned about the extent to which the distribution of instruction hours has an effect on the students' listening, grammar, vocabulary, reading, writing, and speaking skills in English. The second question focuses on whether such effect is the same or different according to the students' initial proficiency level.

Chapter 5 will explain the methodology used for this research project. Section 5.1 will introduce the programs and participants, while section 5.2 will describe the procedures and instruments of data collection. Finally, section 5.3 will include the measures which were selected in order to analyze the learners' performance.

In Chapter 6 the results of the tests performed by the students will be presented. Section 6.1 will report on the results obtained by the students at the intermediate level. First, the descriptive statistics will be introduced, followed by the results of the statistical analyses for the listening, grammar, vocabulary and reading parts of the test (section 6.1.1). Afterwards, the results of the written (section 6.1.2) and oral production tasks (section 6.1.3) will be offered. Section 6.2 will examine the findings from the advanced students with respect to listening, grammar and vocabulary (section 6.2.1), writing skills (section 6.2.2) and oral skills (section 6.2.3). Finally, a comparison will be established between the scores on the written and oral tasks obtained by the students at the intermediate level and the advanced level (section 6.4), as well as by the learners at the advanced level and a group of native English speakers (section 6.5).

The results reported in Chapter 6 will be discussed in Chapter 7. For the

sake of clarity, first, the results achieved by the intermediate-level students will be examined and interpreted in light of similar studies in the SLA literature, as well as theoretical cognitive models (section 7.1). Then, section 7.2 will attempt to provide an explanation for the students' performance at the advanced level. Finally, section 7.3 will include a brief discussion regarding the measures selected in this dissertation to analyze the students' written and oral production.

Chapter 8 will offer a conclusion, in which the findings from this dissertation will be summarized. Then, some ideas will be suggested for further research. The references and the appendix will be introduced after this chapter.

CHAPTER 2: PRACTICE IN SKILL ACQUISITION AND IN SLA

2.1. Practice in skill acquisition

The role of practice in learning or skill acquisition in general has always been considered to be highly relevant; moreover, it has been demonstrated that what is learned is a function of the time devoted to learning (Anderson, 1982; Cooper & Pantle, 1967). Nevertheless, practice or time devoted to learning is not the single factor that affects skill acquisition. Apart from a certain amount of practice, Baddeley (1990) suggests that aspects such as attention, relating new information to previously-known information, as well as consolidation of new knowledge are also crucial when considering what learning is.

The maxim that "practice makes perfect" reflects a common belief that practice is directly related to improvement in performance. Even though it is generally assumed that some people have a certain aptitude to perform specific skills, the role of practice in advanced performance has been emphasized by several authors. Ericsson et al. (1993) claim that what distinguishes expert performers is not a specific innate characteristic that predisposes them to excel in a particular activity, but years and years of effortful, *deliberate practice*. Furthermore, Ericsson & Charness (1994: 725) claim that "[t]he effects of extended deliberate practice are more far-reaching than is commonly believed. Performers can acquire skills that circumvent basic limits on working memory capacity and sequential

processing. Deliberate practice can also lead to anatomical changes resulting from adaptations to intensive physical activity". These authors suggest that expert performance cannot be achieved in less than a decade of deliberate practice in a variety of fields: chess, music, writing, painting, science, etc. Apart from a long period of time devoted to training, according to Ericsson et al. (1993) the characteristics that deliberate practice has are the following: first of all motivation on the part of the learner; then, the tasks used for practice must be well designed, and finally, immediate feedback should be provided.

Extensive time devoted to practice is not only a condition for expert performance; it is also a prerequisite for the acquisition of complex cognitive skills as well as motor skills[3]. Practice is necessary in order to achieve automatization. According to some authors (Baddeley, 1990), working memory capacity is limited; consequently, lower-level processes must be automatized so that working memory resources can be devoted to high-level processing. As Johnson (1996: 137) notices, "the skill of automization is the ability to get things right when no attention is available for getting them right." Before concentrating on the role of practice in skilled performance, a brief review of some issues concerning memory and automatization is in order.

Although there are different theories about human memory, there is a certain agreement that two different concepts should be differentiated, one relating

[3] Even though there are differences in the acquisition and mechanisms required for processing in the case of motor and cognitive skills, both are comparable in terms of the effects of practice in skilled performance. Moreover, even highly complex cognitive skills (such as language acquisition) involve motor actions (lips, tongue, etc.)

to memories that are more permanent and stronger (*long-term memory*, LTM) and another to more active, less durable processes (*short-term memory*, STM or *working memory*, WM)[4]. Early accounts of memory (Atkinson & Shiffrin, 1968) referred to LTM and STM as two storage systems (*long-term store* and *short-term store*). For learning to take place, a subset of the available information has to receive selective attention in sensory registers. Then such information passes to the short-term store (STS). The information that can remain in such store is limited, and, although it can be retrieved effortlessly, due to its high activation, if it is not rehearsed it will be lost. When some information is appropriately rehearsed it then moves to the long-term store (LTS) where it will be permanently registered. The information in LTS is strong, since it is durable (Anderson, 1980); nevertheless, retrieval of such information is more difficult and slower than retrieval from the STS.

Other views on memory include the one presented by Craik & Lockhart (1972), according to whom what differentiates LTM from STM is the level of processing which information undergoes, and not the fact that they are two separate stores. Information that receives deeper processing (in the authors' opinion, semantic processing) will be more durable than information whose processing is more superficial. In fact, more recent accounts of WM emphasize the fact that WM is not only a storage but also a processing mechanism, as opposed to traditional views on memory which highlighted the role of STM as storage.

Other theories of memory (Norman, 1968) propose that there is a

[4] Although some authors claim that WM should be distinguished from STM (Carlson et al., 1990), in cognitive psychology WM and STM are generally considered as two equivalent systems (Anderson, 1983).

continuum between LTM and WM, instead of their being two separate systems. WM is claimed to be a subset of activated representations from LTM; in addition, WM is considered to be the capacity to control and maintain attention.

Notwithstanding the view of LTM and STM as storage mechanisms or processing capacities, attention has always been considered to play an important role in encoding information in STM. According to most models of STM (Baddeley, 1990; Miller, 1956), attention or STM capacity is limited[5]. Complex cognitive skills place high demands on working-memory capacity, and they require a good deal of attention, whether the activity is driving a car or speaking a foreign language. In the former, low-level skills such as changing gears or using a steering-wheel are supposed to be automatized before the driver can devote his/her attention to anticipating what other drivers or pedestrians might do, or to have full awareness of what is happening around him/her. Similarly, in order to speak fluently, an L2 learner needs to have an automatic access to language forms so that he/she can concentrate on the content he/she wants to convey or is being transmitted to him/her.

Automatic behavior in skilled performance has been characterized as fast, efficient, effortless, not hindered by disruption of interfering events, difficult to modify or inhibit, and not limited by working-memory capacity (DeKeyser, 2001; Schmidt, 1992; Segalowitz, 2003; Shiffrin & Schneider, 1977). Some researchers have emphasized that automatization implies more than a quantitative change in performance (speed-up process); automatic behavior implies a qualitative change.

[5] For alternative models of attention see Robinson (1995).

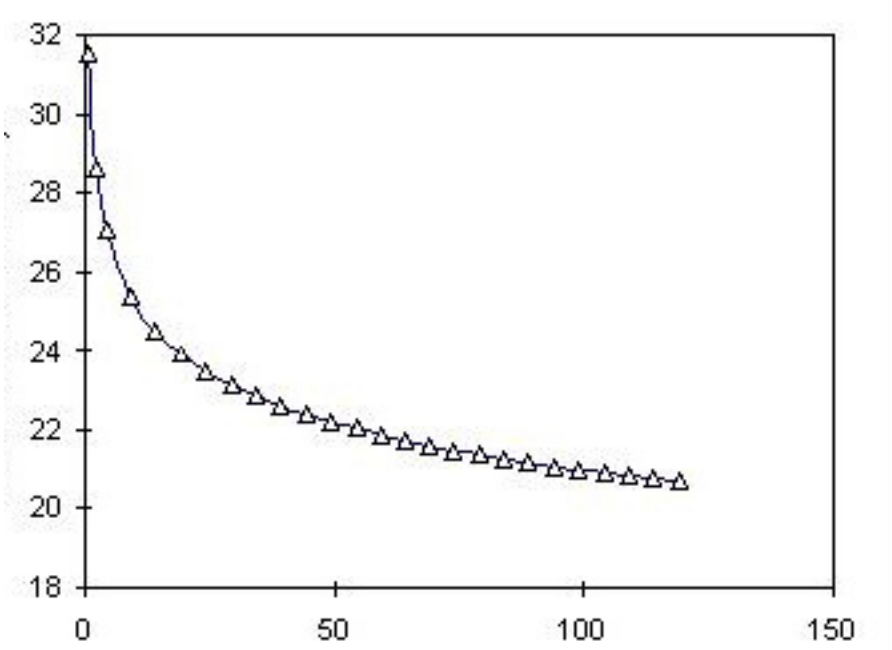
For instance, Ericsson & Charness (1994) suggest that, in the acquisition of new information, improvement in performance through practice (or automatization) is not achieved simply by the capacity subjects attain to keep larger chunks in STM, but by the ability to store information in LTM and thus circumvent the constraints of WM. Cheng (1985) claims that improvement through practice is due to a restructuring of the task components, which implies a change in processing mechanisms, not simply a speed-up process.

Whichever the view on automatization adopted, there is agreement on the fact that massive practice, or extended time-on-task, is a requirement for automatic performance. Newell & Rosenbloom (1981) explained mathematically how practice affects skill acquisition. The authors provide a series of examples regarding the acquisition of perceptual-motor skills, elementary decisions, memory, complex routines, or problem solving, in order to illustrate that, in all these situations, the more often a skill is practiced the more improvement in performance occurs. Newell & Rosenbloom (1981: 2) claim this effect is visible everywhere in psychological behavior, suggesting that there is a ubiquitous law of practice; moreover, practice seems to follow a power law, "that is, plotting the logarithm of the time to perform a task against the logarithm of the trial number always yields a straight line, more or less." Apart from the fact that performance improves with practice, the power law of practice demonstrates that, whereas a considerable progress is experienced at initial stages in the acquisition process, improvement decreases as a function of increasing practice or frequency. The power law of practice is one of the most important characteristics of skill acquisition and the

theories which try to account for the acquisition of cognitive skills should be able to predict this power law (Schmidt, 1992).

Newell & Rosenbloom (1981) and Rosenbloom & Newell (1987) suggest that learning occurs by organizing knowledge in chunks, which vary in terms of length and frequency. The larger a chunk is the fewer possibilities it will have of recurrence. The lower-level chunks are acquired first, since they occur more frequently. Then, as chunks increase in size, their frequency of occurrence decreases; thus the opportunities of practicing high-level chunks are rarer. This fact explains why there is a dramatic improvement as a consequence of practice at early stages of skill acquisition, while progress is less obvious at later stages. Figure 2.1 illustrates the power law of practice, with the y axis representing the time to perform a task and the x axis the number of trials/practice.

Figure 2.1: The Power Law of Practice



Strength theories, such as MacKay's (1982) suggest that the role of practice in skill acquisition is to strengthen the connections between stimulus and response. These theories predict more improvement at initial stages (since it is then that certain connections become stronger), while the effect of additional practice beyond the point that connections are already strong will decrease as a power function of practice.

Another theory which also predicts the power law of practice is Logan's instance theory (Logan, 1988). Automatization occurs, according to Logan, when performance is based on direct memory retrieval of past solutions, as opposed to initial stages in the acquisition process, which depend on algorithm-based (or rule-based) performance. In his experiment, Logan (1988) observed that with increased practice subjects were faster in solving problems such as "A + 4 = ?" (E would be the answer) because, first, learners perform the algorithm, which takes some time. However, with practice, specific solutions are registered in the subjects' memory; consequently, when previously solved problems appear again, the solution can be automatically retrieved from memory, instead of through algorithmic processing. According to the power law, the effects of practice will be more obvious at the stage when subjects switch from algorithmic computation to direct memory retrieval. However, increased practice of memory retrieval will not have such a clear impact on performance.

Logan's instance theory clearly contrasts with Anderson's ACT (Adaptive Control of Thought) theory of human cognition, since Anderson's explanation for automatized performance lies in the transition from instance-based to rule-based

behavior. Anderson's ACT theory, which also uses the power law of practice in order to explain skill acquisition, was designed in the 1970s to account for the acquisition of cognitive skills, by means of the ACT model (Anderson, 1976), ACT* (Anderson, 1983), and finally the ACT-R (Anderson, 1993), which is the last version. Anderson's ACT theory is being updated until the present day (Anderson, Finchman, & Douglass, 1997; Anderson et al., 2004) in its goal to provide an explanation for as many aspects of human cognition as possible. It is practice which, in this model, determines acquisition and automatization of skilled performance.

According to Anderson's ACT-R 5.0[6] (Anderson et al., 2004) the basic architecture of cognition consists of the following systems, which in turn correspond to specific neural associations[7]: the perceptual-motor system (which provides the connection between the external world and cognition), the goal system (associated with working memory), the declarative memory system, and finally the procedural memory system. All of these different modules are independent; however, there is a central production system which coordinates the functions of the different systems. The components which have been given more attention in the ACT theory along its history have been the declarative and the procedural, and it is in those modules where practice has a more important function.

---

[6] The latest version of the ACT theory is ACT-R 6.0, which is available at http://act-r.psy.cmu.edu/actr6/

[7] For a detailed analysis of the brain location of the functions related to declarative and procedural knowledge see Ullman (2001).

Skill acquisition, according to Anderson, implies a transition of knowledge from a declarative to a procedural stage. For Anderson (1993: 18) "[i]ntuitively, declarative knowledge is factual knowledge that people can report or describe, whereas procedural knowledge is knowledge they can only manifest in their performance."

Anderson has claimed that before a skill can be performed, some kind of **declarative** knowledge must be available. In earlier ACT versions (ACT*, Anderson, 1983), retrieval from declarative long-term memory was necessary before performance; in later versions, declarative knowledge needs only be present in working memory before it can be used (Anderson, 1993). Declarative knowledge is added to memory through chunks, which are a means of organizing a set of elements into a long-term memory unit. When chunks are frequently used, they acquire strength and a certain level of activation. According to Anderson et al. (2004: 1042), "the activation of a chunk is a sum of the base-level activation, reflecting its general usefulness in the past, and an associative activation, reflecting its relevance to the current context." Each presentation of a declarative chunk collaborates to its activation according to the power law of practice; that is, more practice/use will lead to increased activation. Conversely, when a chunk is not highly active, its impact decays following a power law of forgetting (Anderson, 1993; Anderson et al., 2004). As a consequence, depending on the level of activation (chunks can only be retrieved if they reach an activation threshold), some declarative knowledge will or will not be available at the time such knowledge is to be used. This part of the theory is highly relevant to the present study, as will be

shown later.

The next stage in skill acquisition is **proceduralization**, in which declarative structures are transformed into procedures for performing a skill (Anderson, 1982). In the same way as declarative knowledge is represented in chunks, procedural knowledge is represented as production rules of the type:

IF the goal is to classify a person

and he is unmarried

THEN classify him as bachelor[8]

The 'IF' part is referred to as the 'condition', whereas the 'THEN' part is called the 'action'. If practice is performed in this direction (condition → action), such practice will have little effect if performance is required in the opposite direction. This phenomenon is known as the 'asymmetry of practice', which predicts little transfer of practice when the knowledge proceduralized in one direction is required for performance in the other. Although, since both uses arise from the same declarative knowledge, some positive effect of practice in one direction could be obtained in performance in the opposite direction (Anderson, 1993). For example, training subjects in performing functions such as (/ 42 7) = ? should improve their ability in that type of computation to the point that automatization is achieved; however, no similar automaticity should be expected if the subject is presented with arithmetic functions such as (/ 42 ?) = 6, despite the fact that both production rules would call on the same declarative knowledge (Anderson & Finchman, 1994).

[8] Example from unit 1 of the ACT-R 6.0 Tutorial (http://act-r.psy.cmu.edu/actr6/).

Production rules, as declarative chunks, acquire strength and activation through practice, and the amount of practice a specific skill receives will determine its automatization (Anderson, 1993). According to the power law of practice, the time to perform a skill will decrease as trials increase. Not only will practice increase speed in performance, it will also enhance reliability. A large improvement will be evident in early stages of skill acquisition, since performance would reflect the compilation of a production rule, which is a more efficient process in performance than retrieving declarative chunks. Then, as production rules are strengthened, more advanced performance will be observed as a consequence of practice, yet the improvement at later acquisition stages will be less obvious (as was also explained by Newell & Rosenbloom, 1981).

Whereas declarative knowledge is quickly acquired, it is 'generative', in the sense that it can be used for different circumstances, and flexible, since it can be easily modified; procedural knowledge is committed to a special use, thus it is not 'generative'. Moreover, procedural knowledge is not so flexible, and altering inaccurate performance is a difficult task. However, an advantage of procedural knowledge is that it is faster, more efficient and does not place high demands on working memory (Anderson, 1993).

Proceduralization is highly important in skill acquisition, due to the fact that retrieving declarative knowledge from long-term memory is tedious and time-consuming, which makes performance slow.

> Interpreting knowledge in a declarative form … has serious costs in terms of time and working memory space. The process is slow because interpreting requires retrieving declarative information from long-term

> memory … The interpretive productions (problem-solving and analogy) require that declarative information be present in working memory and this can place heavy burden on working memory capacity (Anderson, 1983: 228).

Once productions have been compiled and strengthened, with increased practice they become autonomous and faster. Automatization can be described then as "the process of converting declarative knowledge into procedural knowledge, bringing with it all the advantages of the procedural and eliminating all the disadvantages of the declarative" (Johnson, 1996: 90).

Anderson & Finchman (1994) and Anderson et al. (1997) emphasize the importance of examples in the skill proceduralization process. Examples of how a specific rule works are encoded as declarative structure. When presentations similar to the examples appear, and the learner is required to solve a problem that matches the rule in the example, analogy is used to extend the example. However, analogical extension is slow and relatively difficult. With extended practice with examples that instantiate the same rule, the learner begins to encode the abstract relationship between the elements involved in the rule. Such encoding is first declarative, yet, with more practice, production rules begin to emerge (as observed in the directional asymmetry which starts to be observed). Performance following procedural rule use is faster and more effective; nevertheless, it is committed to a specific use and it works in only one direction (as opposed to the declarative encoding of a rule). Still more and more practice and many encounters with the same example may allow direct retrieval of the solution from declarative memory (since the solution, being so strong and active, can be directly retrieved without

application of the production rule), which will lead to a faster and more direct production than the production rule use. Summarizing, Anderson et al. (1997: 945) claim that "performance in a skilled task can reflect a complex mixture of processes. It involves using examples in two ways (analogy and simple retrieval) and using two types of rules (abstract declarative and procedural)."

Even if Anderson remarks that the four stages he proposed for skill acquisition (direct retrieval of the answer from memory, analogical extension, declarative encoding, and procedural rule use) are not strictly sequenced (Anderson et al., 1997); his theory suggests that acquisition occurs as the learner moves from an example-based to a rule-based behavior. Such process is gradual, as opposed to earlier claims made in previous versions of the ACT theory (ACT*). The claim that acquisition consists of an instance-to-rule transition contrasts with Logan's rule-to-instance skill acquisition theory (Logan, 1988).

The implications of the ACT theory can be applied to teaching. Anderson (1993) suggests that a skill should be taught first through declarative instruction, using examples of the production rules. By analogy to those examples, the learner will apply the production rules repeatedly until such behavior is strengthened and reliable. The students' performance must be monitored and appropriate feedback must be provided to make sure that the learner has acquired the right rule and is applying it correctly. For Anderson, time-on-task is the first variable of human learning; therefore, if instruction is to be effective, it should ensure as much time-on-task as it is possible.

Summarizing, time dedicated to practicing a skill is a determining factor

not only in expert performance but also in skill acquisition, according to the research reviewed from cognitive psychology (Anderson, 1993; Anderson et al., 2004; Ericsson et al., 1993; Logan, 1988; Newell & Rosenbloom, 1981). It is only through practice that the declarative knowledge that learners have about a skill can be proceduralized and automatized. Additionally, it is extensive, effortful, deliberate practice which is behind expert performance in many skills (music, painting, writing, etc.) (Ericsson et al., 1993). Apart from the time that practice takes, translated in hours, usually decades of training, there are other aspects of practice which are highly important when analyzing skill acquisition, such as attention during practice, motivation, aptitude, type of practice, feedback, etc. The next section will focus on the effect of time devoted to practice on L2 performance, which is related to the central purpose of this research.

2.2. Practice in SLA

While most authors in the field of cognitive psychology agree that practice plays an important role in automatization (Anderson, 1993; Ellis, 2005; Ericsson et al., 1993; Logan, 1988; Segalowitz, 2000; Segalowitz & Segalowitz, 1993), the notion of 'practice' fell into disfavor for some time in SLA (due to its association with the mechanical drills used by the audiolingual approach) the same as 'intentional learning', 'rehearsal' or 'drills' (DeKeyser, 2007a; Hulstijn, 2001). Similarly, after Chomsky criticized the behavioral skill acquisition model and he emphasized meaningful learning (in which the learner's cognitive system had a central role),

concepts as 'stimulus-response', 'habit formation' or 'automatization' were attacked. Carroll (1981) presents a framework which tries to reconcile cognitive theory and behaviorism by suggesting that language learning (first and especially second) can be said to evolve from a conscious, voluntary choice of responses (which is more in agreement with cognitive theory and differs from behaviorism) to an automatic performance. Carroll (1981: 472) believes that the "acquisition of a language consists of learning pairings of intentive states and linguistic elements, as well as the habits presented in the code component that can make utterances grammatically well formed." For this author, first and second language acquisition follow the same mechanisms; nevertheless, there are some differences with respect to "the kinds of information available to the learner about pairings of cognitive states, linguistic responses and code component habits, as well as in the number of opportunities to apply these parings in order to attain highly automatic response capabilities" (p. 472).

Language acquisition researchers have emphasized that first and second language acquisition are different processes, the former usually leading to master performance, whereas automatized performance (understood as fast and error-free) in the case of the latter is not as common. Several explanations have been presented to account for the differences in L1 and L2 acquisition, among which, the critical period hypothesis has been a popular one (Lenneberg, 1967). The critical period hypothesis states language acquisition cannot occur after a certain time (usually along puberty), due to maturational processes which take place in the brain. L1 acquisition usually happens during the first years of life, where the

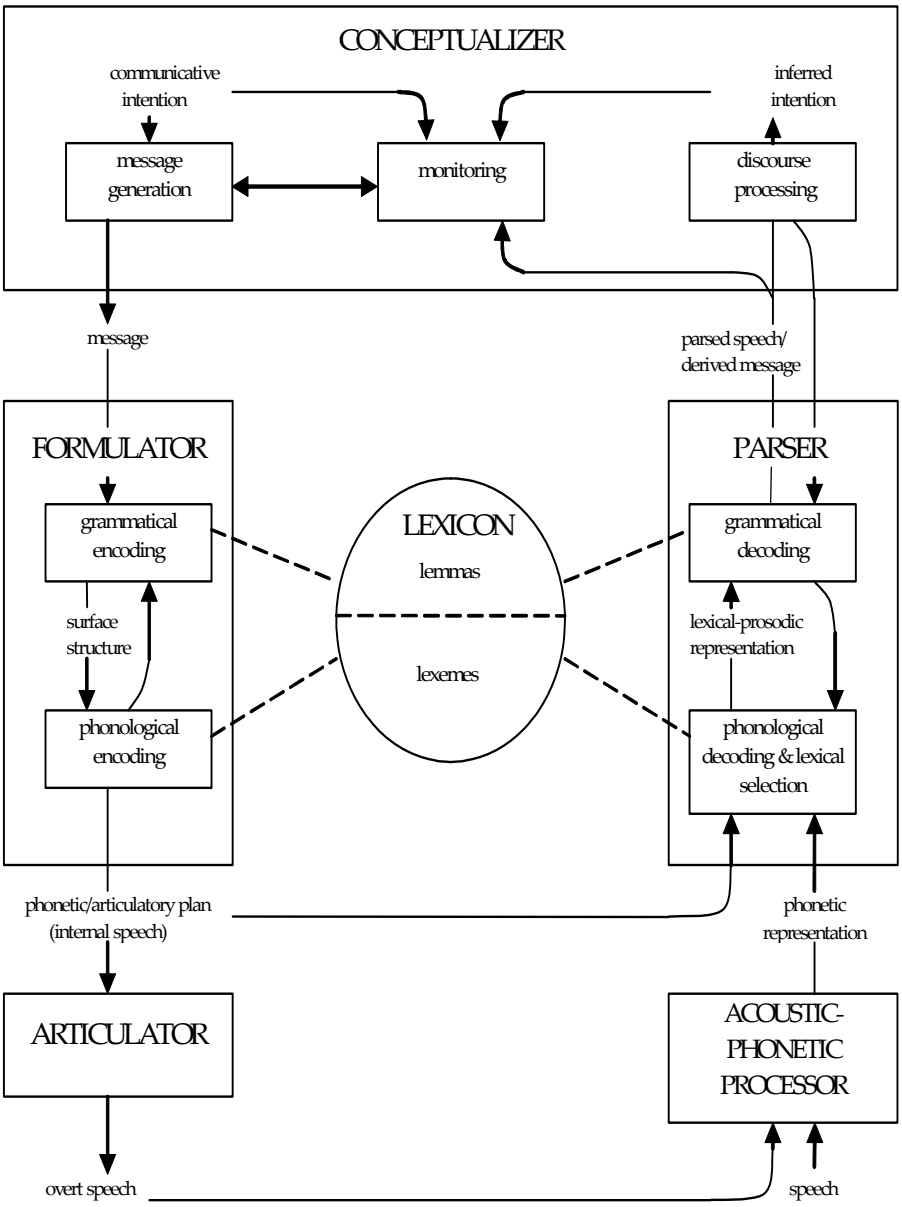
human brain still has plasticity and is ready for the language stimulus. However, SLA tends to occur after such period; therefore, learners hardly ever attain a full command of the L2 (Long, 1990; Singleton, 2005)

Even if this explanation has been widely accepted, not many researchers have considered as an important issue the fact that the time learners spend practicing their native language and the frequency of such practice is drastically superior to the time devoted to learning an L2, and what fluency requires, as N. Ellis (2001: 36) remarks, is a massive amount of hours of practice, of which L2 learners hardly ever dispose: "Fluent language users have had tens of thousands of hours on task. They have processed many millions of utterances involving tens of thousands of types presented as innumerable tokens." Consequently, despite the importance that the age factor can have in SLA in terms of cognitive differences according to the learners' age, considerations of time devoted to practicing a specific language should be considered when discussing language acquisition (Muñoz, 2006).

In the case of L1 performance, according to Levelt (1989, 1993), the speakers need access to two types of knowledge: declarative and procedural. The procedural knowledge has the format "if x then y" (reviewed in Anderson's ACT model in the previous section).

Declarative knowledge mainly comprises encyclopedic knowledge about the world, situational knowledge, or discourse records (which are highly related to the conceptualizer), and also knowledge of the lexicon (lemmas and lexemes). Levelt's model explains both comprehension and production (see Figure 2.2).

Figure 2.2: Levelt's (1993) Language Production Model



In the case of language comprehension, the speech that is in the environment is perceived through the acoustic-phonetic processor. Then, in the parser, phonological decoding, lexical selection and grammatical decoding take place. Finally, the speech moves to the conceptualizer, which checks for success of processing at earlier stages. With respect to production, first, concepts are planned

in the conceptualizer. Next, grammatical and phonological encoding occur in the formulator. Finally, the articulator is in charge of producing the overt speech.

Fluent performance in language production and comprehension will occur when all the above-mentioned procedures are automatized. The more automatized these procedures are, the less need to devote attention to language forms there will be. Levelt argues that mature native speakers use their controlled knowledge in the conceptualizer to construct a specific message. On the other hand, choosing the appropriate grammatical and lexical forms (which takes place in the formulator) occurs automatically. According to Towell & Hawkins (1994: 170):

> The significance of such a model [Levelt's] for SLA is at least twofold. On the comprehension side, fluent comprehension will depend on developing procedures for decoding the message in real-time. On the production side, if language is to be creative (i.e. other than set phrases) the procedures for processing syntax in real-time will have to be developed.

In order for a learner to reach automatization in an L2, many hours of practice are necessary. Apart from a massive amount of time, practice also requires certain characteristics in order for it to be optimal, as suggested in the book edited by Robert DeKeyser (DeKeyser, 2007c). Leow (2007) believes that during practice attention is highly important, and the instructor should focus the students' attention towards the target items in any possible way. Ortega (2007) suggests that for practice to be effective it should be interactive, meaningful and focused on task-essential forms. In addition, *output* practice has been claimed to be of significant importance in proceduralization and automatization of language skills (Muranoi, 2007). There are other authors who have focused on the relationship between

aptitude and practice (Robinson, 2007), age and appropriate SLA practice (Muñoz, 2007), interaction as practice (Mackey, 2007), or feedback during practice (Leeman, 2007).

Even if all the aforementioned aspects related to practice are highly important and should be examined in order to determine what optimal L2 practice is, considering the circumstances which surround a specific learner, the main focus of this particular research is the relationship between time and practice; more specifically, time distribution of practice in SLA (which will be reviewed in the next chapter). There are several authors who have emphasized the importance of practice (understood as time devoted to language learning and opportunities for L2 performance) in order to explain how automatization takes place in second language acquisition (Bialystok, 1994; DeKeyser, 1997; 2007a; 2007b; Ellis, 2005; Hulstijn, 2001; McLaughlin, 1987; 1990; Schmidt, 1992; 2001; Segalowitz, 2000; Segalowitz & Segalowitz, 1993).

In SLA, as in the acquisition of other cognitive skills, learners are claimed to eventually move from a *controlled*, effortful, conscious performance to a more *automatic* performance, which requires less voluntary control and attention on the part of the learner (Bialystok, 1994; McLaughlin, 1987; 1990; Segalowitz, 2000) and resembles more and more the performance of the native speakers. As was mentioned in the previous section regarding the acquisition of complex cognitive skills, humans can only concentrate their attention on a limited amount of information (Baddeley, 1990). In the SLA literature many authors hold this view (Foster & Skehan, 1996; Skehan & Foster, 2001; VanPatten, 1990), suggesting that in

terms of language production or comprehension, learners must prioritize where they allocate their attention: to form or meaning. McLaughlin (1987) argues that native speakers have an automatic access to language forms and generally focus on meaning. Non-native speakers, on the other hand, have to allocate their attention to language forms (which are not completely mastered) as well as meaning, and that is why performance is less fluent. L2 practice will eventually lead to the automatization of language forms and will liberate attentional resources to encode other parts of the message. According to McLaughlin, the acquisition of an L2 is similar to the acquisition of any complex cognitive skill: first, due to limited attentional capacities, performance is highly controlled at initial stages; then, through practice, component skills are automatized so that attentional resources are liberated for high-level processing; finally, a restructuring of the learners' internal representation occurs.

Hulstijn (2001) also suggests that native speakers tend to focus their attention on the concepts conveyed in a specific message, since the automatization of the subskills which are necessary to fluently write, speak, comprehend or read (such as word recognition) has already been completed. As a consequence, what L2 learners need in order to automatize their L2 skills is extensive practice. Hulstijn (2001) claims that automatic access to lexical entries is necessary for communication to proceed efficiently. Such automatization can be achieved through practice, whether it occurs in incidental (implicit) learning conditions or through intentional rehearsal of second language vocabulary.

Similarly, Schmidt (2001: 7) emphasizes the fact that at the beginning stages

of SLA, learners are cognitively overloaded, since they have to pay attention to all the aspects of the L2, "but as simpler processing routines are over-learned, they have more capacity to attend to details, eventually being able to attend to whatever native speakers pay attention to." Bialystok (1994) also agrees that learners' attentional capacities are limited. Fluency or automatization in SLA is the result of practice, through which the learners become more capable of allocating their attentional resources to the relevant areas of the L2[9].

Norman Segalowitz has also investigated the effect of practice on skill automatization. According to Segalowitz (2000), in order for a skill to be automatized in terms of fluent performance (which he describes as performance characterized by fluidity (free of interruptions), high speed and accuracy) a qualitative change in the underlying processing mechanisms in charge of performing the skill in question must occur. Such change is brought about by extensive practice in a way which will be briefly explained below.

Segalowitz (2000) distinguishes two complementary factors which are involved in the execution of skilled performance. The first factor is attention, which monitors performance in order for it to be accurate. The second refers to automatic processing, which is performed in a fast, effortless and accurate fashion despite

[9] As opposed to the previously presented view on attention, Robinson (2001; 2003a) follows alternative models (Wickens, 1989) which suggest that there are multiple resource pools where attention can be allocated (e.g., auditory/visual; vocal/manual), and that there should not be a competition for attention unless specific tasks draw from the same resource pool. Robinson (2001) argues that there are no constraints on attention in SLA as long as L2 tasks do not draw on the same resource pool.

intervening internal sources. According to this author, what differentiates automatization from purely speed-up performance is the qualitative shift, from reliance on control mechanisms (performance fluency) to greater reliance on automatic processing (cognitive fluency). According to Segalowitz (2000) and Segalowitz & Segalowitz (1993), initial practice leads to faster performance, while the underlying processing mechanisms remain unchanged. Further training and practice will little by little make a qualitative change in cognition, with the more inefficient processes disappearing and the stronger patterns becoming automatized. Segalowitz believes that second language skills are automatized in a similar way to other cognitive skills (Segalowitz, 2007), and that first, a speed-up performance can be observed in some second language processing skills (such as word recognition). However, such faster performance is clearly different from automatized performance, in which slower, less efficient processing components are dropped out (Segalowitz, 2000; Segalowitz & Segalowitz, 1993).

Even if Anderson's ACT model reviewed in the previous section has not been generalized to account for second language acquisition, several researchers have shown that Anderson's theory can explain several issues related to SLA, more specifically classroom-based second/foreign language instruction (DeKeyser, 1997; 2007a; Johnson, 1996; O'Malley & Chamot, 1990; Towell & Hawkins, 1994). O'Malley & Chamot (1990) consider Anderson's theory appropriate in order to explain the acquisition of an L2, despite the fact that the authors mention some problems with it. For instance, if declarative knowledge of the grammar rules has to be acquired before the L2 can be practiced, instruction would be highly

frustrating for the students. Generally, learners want to practice their language skills as soon as possible, even if the rules are not thoroughly learned. That is why, waiting until the acquisition of the rules has been completed before allowing the students to start practicing the language in real-life situations could make the students lose their motivation.

O'Malley & Chamot (1990), establishing a parallelism with Anderson's ACT theory, believe that learning strategies should be regarded as one form of procedural knowledge. Inferencing, for example, could be represented in the following terms:

> IF the goal is to comprehend an oral or written text,
> And I am unable to identify a word's meaning,
> THEN I will try to infer the meaning from context.
> (O'Malley & Chamot, 1990: 52)

In his analysis of the effect of practice in the study abroad context, DeKeyser (2007b) also uses Anderson's ACT theory in order to explain why students in such contexts are not usually reported to make many language gains, as compared to their peers who stay at home. According to the author, before learners are immersed in the L2 country, they need to have already gone through the proceduralization stage in their L2 acquisition process; otherwise, they will simply acquire new formulas when faced with communication difficulties in a naturalistic context. These formulas will then be automatized, yet such knowledge will not translate into automatic L2 rule use. Ideally, when the students spend some time abroad, they should have some procedural knowledge of the foreign language rules which, with practice in real-life situations, should become automatized.

DeKeyser (1997; 2007a; 2007b) suggests that classroom instruction should facilitate the transition of knowledge from the declarative to the procedural stage, by providing explicit (declarative) information about the target rule or structure, and then offer ample opportunity for the students to use that declarative knowledge and proceduralize it in challenging and diverse contexts which resemble real-life communicative situations. Finally, when such procedural knowledge is repeated and practiced over a long period of time, automatization is likely to occur.

Moreover, DeKeyser (1997) observed that the effect of practice on comprehension and production of morphosyntactic rules followed the same power function that has been reported for practice in other cognitive skills (Newell & Rosenbloom, 1981). Automatization, as defined by shorter reaction time and lower error rate with respect to both comprehension and production of morphosyntactic rules, was observed in DeKeyser's study as a gradual function of practice. Moreover, in agreement with the power function of practice for other cognitive skills, the results of initial practice appeared different (resulting in higher improvement) from those observed after subsequent practice, since the former instantiate the proceduralization of declarative knowledge, whereas the latter reflect gradual automatization (DeKeyser, 1997). The second main finding reported by DeKeyser supports the skill specificity of procedural practice proposed by Anderson in the sense that those students who practiced comprehension and improved their performance in that skill did not show equivalent performance in production. Similarly, those learners who practiced production did not evidence

transfer of such skill in comprehension-based tasks. Examining the results of his study, DeKeyser (1997: 211) claims that:

> [P]ractice in the use of second language grammar rules, whether through comprehension or production tasks, has exactly the same effect on learning as it does in the acquisition of other cognitive skills. This strongly suggests that the same mechanisms are involved and that the ability to comprehend or produce sentences in a second language is not necessarily acquired through the implicit mechanisms of a separate mental module.

Consequently, according to this author, second language acquisition can be explained through the same mechanisms as the acquisition of other cognitive skills, as presented in Anderson's ACT model. More specifically, second language learners initially have a declarative knowledge of the L2, which, through practice, can turn into qualitatively different procedural knowledge. After proceduralization has occurred, additional practice will lead to a slower process of gradual automatization.

Ranta & Lyster (2007) adapt the stages of learning suggested by Anderson (1983) to their context of research: French immersion programs in Canada. The three stages of skill acquisition proposed in Anderson (1983) are the following: cognitive (which involves conscious effort on the part of the learner and mainly consists of learning of rules and acquisition of factual knowledge), associative (through practice, procedural knowledge begins to be acquired, although performance is still slow and may evidence some mistakes), and finally autonomous (which is characterized by automatized performance). Ranta & Lyster (2007) propose a sequence for the acquisition of oral skills in French immersion in order to enhance students' performance not only in terms of fluency but also of

accuracy. The proposed sequence starts with an *awareness* stage (similar to Anderson's declarative stage), where the learners pay attention to the target features and notice the gap between their use of the L2 and the corresponding target forms. Then, there is the *practice* phase, in which, through communicative drills the students practice their declarative knowledge in order to achieve proceduralization. Finally, during the *feedback* phase practice is oriented towards communication and is less controlled. In this phase, corrective feedback leads to modified output, which is considered as practice that also leads to proceduralization of declarative knowledge.

Another researcher who has studied how practice affects SLA, this time from a connectionist perspective, is Nick Ellis. Ellis (2005) agrees with the maxim "practice makes perfect" for skill acquisition in general and for language acquisition in particular. This author has argued that language acquisition takes place implicitly from usage. Nevertheless, explicit learning also has an important role in language acquisition, especially in second language learning. Before explaining in more detail the role of practice in Ellis' acquisition model, it is necessary to review briefly Ellis' approach to language acquisition.

According to Ellis (2005: 306), "[m]ost knowledge is tacit knowledge; most learning is implicit; the vast majority of our cognitive processing is unconscious." This author, as the ones mentioned in this section, believes that the acquisition of language occurs through the same processing mechanisms that are involved in the acquisition of other cognitive skills. Human beings are claimed to learn by association; in the case of language acquisition, it is the form-meaning mapping

which the learner needs to make. Input frequency[10] will determine the acquisition of specific form-function associations. Ellis (2002) analyzes the role of frequency in the processing of phonology, phonotactics, reading, spelling, lexis, morphosyntax, formulaic language, language comprehension, grammaticality, sentence production and syntax. Ellis (2002: 143) claims that language acquisition "is the piecemeal learning of many thousands of constructions and the frequency-biased abstraction of regularities within them." In that sense, language acquisition is implicit (unconscious) learning, in which frequency of usage determines the availability of specific patterns. Moreover, it is through implicit learning that different components are associated (or chunked together) in larger formulas and organizational structures. Implicit learning thus needs thousands of hours on task (practice) in naturalistic environments where language learners are exposed to a vast amount of input, and have a large number of opportunities for output practice.

Nevertheless, Ellis also emphasizes the importance of explicit (conscious) learning in language acquisition, especially in the case of SLA. Explicit learning in SLA can be involved in the initial registration of patterns (or formulas) which are usually highly frequent, salient, prototypical and highly functional (Ellis, 2005). Once several form-meaning exemplars have been learned explicitly, it is through implicit learning that they are bound together; and as a function of frequent co-occurrence in use, their strength increases (for this processing explicit learning is

[10] Such frequency is a function of *type* frequency in the case of phonological, morphological and syntactic patterns, and a function of *token* frequency for irregular forms and idioms (Ellis, 2002).

no longer necessary). However, explicit learning still has an important function, which is changing non-target-like automatized behaviors. Patterns in the L2 which are not salient or are redundant (e.g. 3rd person singular *–s*) may remain unnoticed by the learner, as well as structures which are different from the L1. Those patterns will therefore be automatized in a form that is not target-like, and in order to make the learner notice the target features explicit instruction is necessary. Additionally, explicit memories of utterances can be subsequently used through analogy in creative production of similar utterances.

Consequently, both, conscious and unconscious processes occur in language acquisition, even though implicit mechanisms are considered more powerful. Practice has a highly important role in automatization of language patterns. Ellis (2005) discusses four different effects of practice, two of which are more related to the elaboration of explicit/declarative knowledge, whereas two other effects of practice have more impact on the proceduralization and automatization of explicit knowledge.

The first effect of practice is improved access: the more often declarative/explicit memories are retrieved, the most efficient such retrieval will be. The second effect of practice is schematization and script-building: the sequences that are repeated are constructed more and more skillfully as more practice occurs. The knowledge about those sequences will still be explicit; therefore, through practice, the use of declarative knowledge becomes faster, yet it is not automatized nor the product of implicit competence[11]. Through more

[11] This distinction between fast declarative knowledge and automatization is also made by

practice and through implicit learning, specific utterances are chunked together into entrenched formulas which can be retrieved as such from memory when they are needed. Finally, more implicit practice will lead to automatization of production, which will no longer be under explicit control and will not require attentional resources.

Additionally, Ellis also discusses the power law of practice in order to explain some frequency effects. First of all, Ellis (2002: 152) states that frequency effects are stronger at earlier stages:

> [T]he effects of practice are greatest at early stages of learning, but they eventually reach asymptote. We may not be counting the words as we listen or speak, but each time we process one there is a reduction in processing time that marks this practice increment, and thus the perceptual and motor systems become tuned by the experience of a particular language.

Moreover, following the power law of practice, Ellis suggests that the reason why frequency effects are not strong in the case of regular patterns is that performance on those is close to asymptote. That is, whereas it has been shown that for irregular verbs frequency of use determines speed of processing, in the case of regular verbs such a distinction between frequent and non-frequent items does not seem to have an impact on processing.

Increases in practice opportunities may also naturally involve increases in time, which, in the case of L2 instruction is claimed to lead to higher language proficiency (Stern, 1985). Whereas there is agreement that increases in time have a

---

Segalowitz & Segalowitz (1993).

positive effect on L2 learning, the way in which the additional hours of instruction should be distributed is controversial: "Time allocations have been, and frequently still are today, a matter of tradition and guesswork" (Stern, 1985: 18). Some programs have increased the hours of foreign language instruction by offering more concentrated language exposure without adding extra weeks or extra years to the arranged period of foreign language learning (intensive French or intensive English classes in primary schools in Canada). Other alternatives include the introduction of the foreign language at earlier grades in school, thus increasing the period of language teaching while reducing the amount of contact with the foreign language each week (Muñoz, 2006). Usually, the models that increase and concentrate language instruction have proven to be more successful (Collins et al., 1999). The next chapter will concentrate on the effects of time distribution on learning in general and more specifically on L2 learning.

2.3. Summary of Chapter 2

Chapter 2 has analyzed the role of practice, understood as opportunities to perform a given skill, in the acquisition of cognitive skills in general, and in SLA in particular. It has been claimed that in order to attain automatized performance in any complex cognitive skill, massive practice is necessary. The results of such practice will be more clearly observed at the beginning stages of the acquisition process than at later stages, according to the power law of practice. Automatization is required for skillful performance due to the fact that humans have limited

attentional capacity. Attention should be devoted to high-level processing; consequently, low-level processes need to be automatized for fluent and reliable performance.

In SLA practice is also necessary for L2 skills to become automatized. At beginning stages in the acquisition process the learners need to pay attention to all the aspects of the L2. Gradually, as components of different skills are automatized (e.g. lexical access for oral production), learners can allocate their attentional resources to focus on what usually native speakers pay attention to: meaning rather than language forms. Prior to the automatization stage, long periods of time-on-task are necessary before the knowledge that L2 learners usually obtain in a declarative form becomes proceduralized and used automatically, as demonstrated by a fast and error-free performance.

CHAPTER 3: TIME DISTRIBUTION OF PRACTICE: MASSED VS. DISTRIBUTED LEARNING

Among the different factors that affect memory, repetition is one of the most powerful (Hintzman, 1976). That repetition enhances memory and retention is an undisputed fact; nevertheless, several researchers have demonstrated that depending on how repetitions are distributed, the effects on memory and retention will differ (Dempster, 1996)[12]. Many experiments have been performed in order to analyze in which way the repetitions of items to be learned should be presented, whether in a distributed form (with some intervening items in between, or with some time in-between learning sessions) or in a massed form (with repetitions following each other in a concentrated fashion). Apart from laboratory experiments conducted by cognitive psychologists, there have been several proposals for education programs that have tried to address the issue of time distribution in actual teaching in schools or universities. Different reforms have been applied in the school curriculum in middle schools or high schools (the most popular one being the block scheduling), as well as in different colleges (through intensive or accelerated courses).

This chapter will present, first of all, some studies that have investigated how distribution of to-be-learned items affects memorization and retrieval of those

[12] There are other aspects that influence the effectiveness of repetitions on learning apart from time distribution, such as meaningfulness of repetitions, strategies used during learning episodes, attention, motivation to learn, etc.

items. Then, several educational reforms dealing with time distribution will be mentioned for subjects other than languages. Finally, research on time distribution of foreign language instruction hours will be reviewed.

3.1. Massed and distributed learning: Evidence from the cognitive psychology literature

In the cognitive psychology literature, study conditions in which repetitions of items to be learned appear in spaced sequences have been found to be more favorable for long-term retention than presentations in which repetitions occur instantly. This has been a long-standing claim, as can be observed in two quotes presented by Dempster (1996), one from Ebbinghaus (1885: 1913): "with any considerable number of repetitions a suitable distribution of them over a space of time is decidedly more advantageous than the massing of them at a single time", and the other from McGeoch (1943: 140) "if two associations are of equal strength but of different age, a new repetition has greater value for the older one". Abundant evidence has been provided for the positive effect of spaced sequences on learning by a vast number of studies, especially during the 1970s and 1980s (Cuddy & Jacoby, 1982; Dempster, 1987; Glenberg, 1979; Greene, 1989; Hintzman, 1976; Melton, 1970; Toppino & DiGeorge, 1984), but also in more recent times (Appleton et al., 2005; Braun & Rubin, 1998; Delaney & Knowles, 2005; Donovan & Radosevich, 1999; Mammarella, Russo, & Avons, 2002; Pavlik & Anderson, 2005; Raaijmakers, 2003; Seabrook, Brown, & Solity, 2005; Verkoeijen, Rikers, & Schmidt,

2005). The phenomenon under analysis has been called the *spacing effect*, according to which including intervals between learning episodes (distributed practice) is more effective for subsequent memorization and retention than 'massed' training episodes, in which learning takes place in a concentrated period of time. The following verse by the psychologist Ulric Neisser quoted by Bjork (1988) is highly illustrative in terms of the belief about the superiority of distributed over massed practice:

You can get a good deal from rehearsal
If it just has the proper dispersal.
You would just be an ass
To do it *en masse*:
your remembering would turn out much worsal.

Although the spacing effect has been found mostly in memorization of words that appear in a list (whether they are in the participants' native language or in a foreign language), spacing effects have also been obtained in many other types of tasks, such as remembering advertisements (Appelton et al., 2005), remembering unfamiliar faces (Mammarella et al., 2002), acquisition of certain mathematical skills (Gay, 1974), or text processing tasks (Glover & Corkill, 1987; Krug, Davis, & Glover, 1990). Since the spacing effect has been found in so many different types of tasks, it has been claimed to be a fundamental aspect of learning and memory.

Although the spacing effect is found constantly in most research contexts within the psychological literature, there is not much agreement as to why immediate repetitions of an item are not as effectively recalled as spaced repetitions. Several theories have been proposed, most of them falling into the following groups: encoding variability theories, deficient-processing theories, or

study-phase retrieval accounts, although lately many authors use at least two of these different accounts to present a more precise explanation for the spacing effect (Braun & Rubin, 1998; Raaijmakers, 2003; Riches, Tomasello & Conti-Ramsden, 2005; Russo & Mammarella, 2002; Verkoeijen et al., 2005). Moreover, recently, there have been mathematical models which have been proposed to account for the spacing effect, as those provided by Raaijmakers (2003) or Pavlik & Anderson (2005).

Encoding variability theories emphasize the fact that spaced items are better recalled than massed, because each presentation in the former distribution is encoded differently, thus providing more retrieval cues. Some encoding variability theories highlight the role of the context (*contextual variability theories*) and claim that the context in which an item appears is encoded together with its meaning (Anderson & Bower, 1972). Melton (1970) suggests that distributed practice allows the storage of more different context cues than massed practice, since the context in which the items occur in distributed presentations is more diverse than when items immediately follow each other. As a result, spaced sequences are better retrieved than massed due to the fact that there are more available contextual cues to help such retrieval. According to Glenberg (1979), encoding variability can account for the spacing effect in free-recall tasks, since recall highly depends on contextual cues, which are claimed to be automatically encoded.

McFarland, Rhodes, & Frey (1979) present an alternative encoding variability theory, suggesting that the spacing effect occurs because *different semantic features* of an item are activated when P1 (first presentation) and P2

(second presentation) are widely spaced, but not when they appear in a massed distribution. As stated in McFarland et al. (1979: 163), "the semantic reading given an item at P1 will be the same as that assigned at P2 if the repetitions are massed; however, if the repetitions are distributed, the interpretations are likely to differ". As a consequence, subjects recall spaced repetitions more effectively than massed repetitions. McFarland et al. (1979) show that the spacing effect is attenuated when massed repetitions highlight different semantic features.

Greene (1989) supports the appropriateness of the encoding variability approach in free recall, yet he suggests that the spacing effect cannot be accounted for exclusively on the basis of this theory (he presents findings on cued-memory tasks as an illustration[13]). Other researchers have also stressed the fact that contextual variability alone cannot account for the spacing effect (Challis, 1993; Postman & Knecht, 1983; Russo et al., 1998).

Deficient processing theories (Challis, 1993; Greeno, 1970; Hintzman, 1976; Jacoby, 1978) claim that the second presentation of massed items does not receive enough processing, because the previous presentation is still too recent. In contrast, when an item is presented after some time has elapsed and after some intervening items have been shown, full processing will be necessary, since the previous

[13] In order to test for learning, different activities can be used to retrieve the material that has been presented. The test can involve recognition (subjects say whether item X has been presented before or not) or recall. In free recall, subjects are asked to provide as many items as they can remember with no cues to help retrieval; conversely, they may be asked to recall material with the help of cues (e.g., what's the word in Spanish for *spring*?) Usually, recognition is an easier task than recall; moreover, cued-recall facilitates retrieval more than free recall.

presentation will not be as readily available as in the case of massed sequences. Cuddy & Jacoby (1982: 465) explain the spacing effect in the following manner: "If the trace of a prior presentation is too readily accessible when an item is repeated, few of the operations originally required to encode that item will be repeated and the result will be an impoverished trace of the later presentation."

Some deficient-processing accounts emphasize that inattention is voluntary on the part of the subject. On the other hand, other authors have dismissed this fact, arguing that such deficient processing can only occur automatically. Dempster (1987) claimed that the spacing effect takes place because subjects *voluntarily* decide to pay less attention to the repetition of a massed item, since they consider that not much attention is necessary due to the recency of the first presentation. Similarly, Rundus (1971) considers that P2 is not completely processed when it appears shortly after P1 because the subject decides not to devote much time to the rehearsal of the second presentation of a massed item. Greene (1989) also underlined the importance of rehearsal in cued-memory tasks, since, in his experiments, no spacing effect was found for incidentally learned material. This finding contrasts with the one reported by Challis (1993), whose study indicated that the spacing effect was produced in both incidental and intentional study conditions. This author concludes that the spacing effect depends on the level of processing more than on the intentionality of the learner, with semantic study conditions (whether they are incidental or intentional) producing spacing effects, whereas graphemic study conditions do not produce such spacing effects.

An explanation of the spacing effect which relies on voluntary strategies

has always been debatable for a number of reasons. First of all, the spacing effect has been found in incidental learning (Challis, 1993); additionally, it has also been found in very young children (Seabrook et al., 2005; Toppino, 1991); finally, the ubiquity of the spacing effect cannot be easily accounted for solely by resorting to voluntary rehearsal or attention on the part of the subjects.

Other deficient-processing theories, instead, consider the mechanisms underlying the spacing effect as *involuntary or automatic*, associated with the basic neurophysiology of memory. Landauer (1969), with his consolidation hypothesis, suggests that P2 in massed sequences interrupts the processing of P1, which is still in a transient state in memory; as a consequence, this interruption prevents consolidation. Contrarily, for spaced presentations P2 occurs after P1 has moved to a more permanent state in memory; therefore, spaced items will be more consolidated than massed items. Similarly, the habituation hypothesis presented by Hintzman, Summer, & Block (1975) maintains that when a presentation is being encoded, some time is necessary before that encoding mechanism can respond to a second presentation. This process accounts for the fact that spaced presentations are better encoded than massed, since the former allow more time for recovery of the first presentation to take place.

A similar model is the one proposed by Glanzer & Duarte (1971), the 'Two-storage model of recall'. According to this model, for an item to be fully processed it needs to be registered in the long-term store (LTS). When an item is presented for the first time it stays in the short-term store (STS), and if the second occurrence of that item takes place when the memory of the first one is still in the STS, it is

ignored (this is what Glanzer & Duarte refer to as 'cancellation of duplicates'). On the other hand, if the first presentation has already left the STS, the item is more likely to transfer to the LTS.

Within these deficient-processing theories, there are a large number of studies which have suggested that accessibility of previous encodings may be essential in order to explain the spacing effect (Cuddy & Jacoby, 1982; Dellarosa & Bourne, 1985; Dempster, 1988; Glover & Corkill, 1987; Jacoby, 1978; Krug et al., 1990; Whitten & Bjork, 1977). According to accessibility theories, massed repetitions are not fully processed because previous encodings are still accessible, which is not the case for spaced items. As a consequence, spaced items are better recalled than massed, since longer intervals make previous encodings less accessible, thus more processing is necessary. Similar claims have been made in more recent studies, which emphasize the fact that the less accessible previous presentations of an item are when additional presentations occur (whether accessibility is hindered by time, intervening material or other conditions), the more beneficial such repetitions would be for long-term retention (Pavlik & Anderson, 2005).

Some evidence for the claim that difficulty of accessibility of previous encodings aids subsequent recall is provided by some studies in which paraphrased rather than verbatim repetitions of massed material were included. In such situations it was not time which made difficult the accessibility of P1 at the time of P2 as in most experiments analyzing the spacing effect, but different formulation of the material. As a consequence, even if P1 and P2 appeared within a

short time interval, the accessibility of P1 at P2 was hindered by a modification of the material being presented. Glover & Corkill (1987) found that massed repetitions of brief lectures were better remembered when they were paraphrased instead of verbatim. Furthermore, recall for paraphrased massed repetitions was as effective as for spaced paraphrased repetitions. These results are in agreement with Dellarosa & Bourne (1985: 533), who argued that "anything that increases the probability of a repetition receiving full processing, or conversely, anything that decreases the probability of the item being recognized as a repetition, should improve memorability of the item." Although this theory about accessibility of previous presentations has accounted for many instances of the spacing effect, it must be born in mind that when accessibility of P1 is not simply hindered but made impossible due to very long lags between P1 and P2[14], contrary effects can be obtained. If the trace of P1 is no longer in the participant's memory, P2 will not be able to strengthen that trace (Verkoeijen et al., 2005) and the repeated presentations may simply count as two separate items.

An issue at stake for deficient-processing accounts of the spacing effect is the kind of processing that presentations undergo. According to Challis (1993) the spacing effect is due to different semantic processing (which can be voluntary or involuntary) for massed and spaced presentations. Semantic processing for massed items is weaker than for spaced items, because the first presentation primes its immediate repetition. Therefore, for Challis (1993) semantic priming underlies the

[14] What is considered 'long lag' differs according to the study. In most cognitive psychology experiments the period may vary from a few seconds to a few minutes between different presentations of the same item.

spacing effect in cued-memory tasks. In contrast, Mammarella et al. (2002) also found the spacing effect in cases where no semantic processing was possible (non-words or unfamiliar faces as targets), claiming then that short-term priming, whether it is semantic or non-semantic, underlies the spacing effect.

Other accounts of the spacing effect also include **study-phase retrieval theories**, according to which retrieval of the first presentation when the second one occurs is essential for better recall (Appleton et al., 2005; Braun & Rubin, 1998; Russo, Mammarella, & Avons, 2002; Thios & D'Agostino, 1976; Toppino & Bloom, 2002; Toppino, Hara, & Hackman, 2002). According to Thios & D'Agostino (1976), an important role of the second presentation is to serve as cue to activate retrieval of old information. These researchers observed in their experiments that when subjects were asked to retrieve the first presentation at the time of the second, a spacing effect took place in subsequent free-recall tasks. However, when subjects were simply given the second presentation, without them needing to retrieve the first one, no significant spacing effect was obtained. These results could not be explained using encoding variability or deficient-processing theories; hence, the researchers suggested that the spacing effect depended on study-phase retrieval. Furthermore, Braun & Rubin (1998) claim that time in working memory is an important factor to be considered in study-phase retrieval. When P2 appears some time after P1, P1 stays in working memory longer before it is retrieved at P2 than when P2 appears immediately after P1, which is why spaced items spend more time in working memory and are thus better recalled than massed material. The finding that too large spacings can have negative effects on memory is congruent

with results from other researchers (Gay, 1973; Toppino et al., 2002).

Toppino & Bloom (2002) also claimed that a study-phase retrieval mechanism provides a plausible explanation for the spacing effect. They found in their experiment that items in fast-rate presentations (3 seconds per item) were significantly better recalled in free-recall tasks than items which were slowly presented (10 seconds per item). Consequently, they claimed that successful study-phase retrieval decreases with increased spacing, since considerable spacing would make the retrieval of the first presentation more difficult; therefore, "the point of maximum performance would be expected to occur at higher or lower levels of spacing depending on whether conditions were more or less favorable for successful study-phase retrieval" (Toppino & Bloom, 2002: 443). Similarly, Bahrick & Phelps (1987: 349) claim that "the optimum interval is likely to be the longest interval that avoids retrieval failures."

As previously stated, many researchers, especially in recent times, try to account for the spacing effect using a variety of the theories which have been described, most of them including study-phase retrieval (Appleton et al., 2005; Pavlik, 2005; Pavlik & Anderson, 2005; Raaijmakers, 2003; Verkoeijen et al., 2004; 2005). Appleton et al. (2005) combine study-phase retrieval with the accessibility theories mentioned before, claiming that retrieving P1 at the time of P2 strengthens the memory trace of the item; however, the harder such retrieval process is, the stronger the memory trace will be.

The model presented by Verkoeijen et al. (2004; 2005) is highly comprehensive, and provides an explanation of the spacing effect which includes

contextual variability and study-phase retrieval. According to contextual accounts, the context in which presentations occur is automatically stored and facilitates retrieval of information, especially in cases in which P1 and P2 are in spaced sequences, since the context is more diverse. However, the study-phase component would dictate that contextual change from P1 to P2 is only stored if P1 is retrieved at the time P2 appears. Spacing effects are expected to occur only for material which has undergone successful retrieval. Taking into account the two theories, the spacing effect would be explained by two opposing processes. On the one hand, with increasing spacing, there is more context variation, which is supposed to help retrieval, but on the other, if the space separating P1 and P2 is too large, P1 will not be retrieved and contextual elements will fail to be incorporated, as predicted by the study-phase retrieval theory. According to Verkoeijen et al. (2005), spacing the presentations of two items will initially improve memory, until P1 and P2 are so widely spaced that P1 is not retrieved at P2.

Raaijmakers' SAM (Search of Associative Memory Model) (Raaijmakers, 2003) model considers two mechanisms that cause spacing effects: one is failure for massed items to be encoded (when P2 appears P1 is still in short-term store and thus P2 does not get a new encoding, which is a similar claim as the one made by deficient-processing theories), and the other is contextual overlap in massed sequences (accounted for by contextual variability theories). Raaijmakers (2003) presents in mathematical terms the formulas that would explain the spacing effect taking into account the two mechanisms considered.

Pavlik (2005) and Pavlik & Anderson (2005) provide another mathematical

model of the spacing effect. The model attempts to explain the results obtained in their experiments, in which participants were asked to learn Japanese-English word pairs distributed in massed or spaced sequences. After the training sessions, the participants were then tested at different intervals. Their findings were the following: 1) performance improves with practice (the participants' scores became better across trials, whether in spaced or massed sequences); 2) with delay, there is a decrease in performance (at short retention intervals massed sequences were better recalled, because they were more recent); 3) spacing items is beneficial when test occurs with greater delay (at long retention intervals spaced items were better recalled than massed); 4) the greater the delay, the more the advantage of spaced over massed sequences (the longer the retention interval was, the better spaced items were recalled with respect to massed); 5) the more practice trials there are, the more spacing is important (the benefit of spacing increased as there were more practice trials). Consequently, this model attempts to explain the effects of practice, and the conditions for short-term learning and retention. According to Pavlik & Anderson (2005), each time an item is practiced it is strengthened, but these increments of strength decay as a power function of time. The rate of decay will be determined by the degree of activation at the time of the presentation. In massed sequences, activation of P1 at the time of P2 is high, thus preventing encoding of P1 into long-term memory. When activation of P1 is weak at P2, long-term memory encoding occurs. As a consequence, the higher the degree of activation at the time of the presentation, the faster the decay rate will be, or, in other words, the shorter time that presentation will be retained. Additionally, Pavlik & Anderson (2005)

provide a neurological explanation for the fact that massed items are not as well processed as spaced: long term potentiation does not occur as strongly for massed items because the neurons do not get as much time to recover after a presentation appears.

The mechanisms described by Pavlik & Anderson (2005) explain why massed items may be better recalled at short retention intervals. That is, the authors describe why learning in massed sequences is faster: the presentations are temporarily more increased in activation, due to recency effects; however, since those items are not encoded in long-term memory, they are not so well recalled when the retention interval increases, in contrast with spaced items, the learning of which is slower but long lasting. The claim that two closely spaced presentations will not be fully processed (as opposed to two presentations for which P2 is not perceived as repetition of P1) has already been made by the deficient-processing accounts of the spacing effect presented previously (Dellarosa & Bourne, 1985). Pavlik & Anderson's model explains mathematically what other researchers had previously found: that at short retention intervals massed material is better recalled than spaced, but spaced items are better remembered at long retention intervals. Similarly, Austin (1921) (as reported in Dempster, 1988) found that spaced readings were much more effective than massed in delayed tests, even if in tests of immediate recall there was no difference between massed or spaced readings. Likewise, Bahrick & Phelps (1987) show that students who learned foreign vocabulary in spaced sessions retained better those words after eight years than those students who learned the words through massed practice. Bloom &

Shuell (1981) also demonstrate that performance in a delayed test of vocabulary (four days after the completion of their study) was better for students who had followed distributed practice than for those who had learned under massed conditions. Similarly, Spieler & Balota (1996) report that in their experiment there was a significant interaction between spacing repetitions and test delay, with massed repetitions being better recalled at the immediate test and spaced repetitions at the delayed test. That spaced learning leads to an enhanced retention has not only been demonstrated in the case of vocabulary learning, but also in the case of more complex skills, such as learning of mathematical rules (Gay, 1974).

Although most research on the spacing effect has been oriented towards explaining why it occurs, using one or several of the previously presented theories (contextual variability, deficient-processing or study-phase retrieval), some researchers have emphasized that a variety of factors must be considered in order to explain the spacing effect. Delaney & Knowles (2005) claim that the strategies used in rehearsing are determining in later recall. The authors found that spacing effects took place when participants used deep, elaborative strategies (story mnemonics) but not when they used shallow encoding strategies (rote rehearsal) in unmixed lists (lists which only had spaced or massed items, but not both). Elmes, Craig, & Herdelin (1983) suggest that 'affect' plays an important role in the spacing effect, since the more 'affective' the reactions are towards the words to be learned (whether such reactions are positive or negative) the better later recall will be. Spacing effects are eliminated when either very pleasant or unpleasant words are included.

There has been a debate about whether the spacing effect develops with age and whether young children's learning is also affected by the distribution of to-be-learned material. According to Wilson (1976), the spacing effect is related to working memory since those subjects who had a higher working memory capacity in his experiment (older children) seemed to benefit more from repetitions at longer lags than did younger children (whose working memory capacity is lower). Consequently, Wilson (1976) suggests that the spacing effect develops with age. Toppino & DiGeorge (1984) supported Wilson's thesis illustrating that a spacing effect occurred in picture recognition for first graders, but not for preschoolers, which led them to conclude that the spacing effect may not be hard-wired into our memory system (as some encoding variability theories suggest). Instead, it could be due to a strategy used by subjects by which, voluntarily, they do not pay much attention to the second presentation of an item in a massed distribution (more in agreement with deficient processing views); moreover, this strategy is assumed to develop with age (between four and six years).

Toppino & DeMesquita (1984) further elaborate this idea that the spacing effect develops with age, claiming that neither encoding variability hypotheses nor deficient-processing theories can fully account for their data, which shows different spacing effects in the case of elementary school children. As a consequence, they suggest that a developmental approach to the study of the spacing effect should be pursued in order to explain why the spacing effect seems to be more obvious for older children and adults than for younger children.

In contrast, Seabrook et al. (2005) discovered in their experiments that

spacing repetitions was effective for learning not only for adults and older children but also for younger children. Moreover, the researchers did not find different spacing effects for children of different ages and adults; thus, they ruled out the explanation that the spacing effect develops with age (Toppino & DeMesquita, 1984; Toppino & DiGeorge, 1984; Wilson, 1976) or that the spacing effect is related to working memory (Wilson, 1976) or organizational strategies, as claimed by Toppino & DeMesquita (1984). The results reported by Seabrook, et al. (2005) are however consistent with claims made by Spieler & Balota (1996), who suggest that the spacing effect, and the interaction between spacing of repetitions and test delay, is a central aspect of learning and memory disregarding age.

As it can be seen, there is not one single explanation for why spaced repetitions are better recalled than massed repetitions, despite the fact that the spacing effect is a robust phenomenon. Several researchers have pointed out that deficient processing theories or contextual variability accounts are not sufficient by themselves to account for the spacing effect. Greene (1989) claimed that in order to describe spacing effects in free recalls, contextual variability theories can be used; however, deficient processing provides a better explanation for cued-recall.

Despite the fact that the spacing effect seems to be one of the most reliable, robust and ubiquitous phenomena in cognitive psychology (Dempster, 1988: 627) it has not been usually applied to classroom learning. Dempster (1988) lists some impediments to its application, the most important being that, at the time, the spacing effect had not been demonstrated satisfactorily in school-like activities, since most experiments analyzing spacing effects took place in laboratories.

Nevertheless, Dempster (1988; 1996) considers that the spacing effect has some potential for classroom learning. Moreover, recently, some researchers have tried to take the experiments in the psychological literature from the laboratory to the classroom.

Seabrook et al. (2005), in addition to performing two experiments in a laboratory, also included an experiment in a classroom setting, analyzing the effects of teaching literacy in 'clustered' sessions (which are claimed to be similar to massed) and distributed sessions. The children who followed distributed teaching lessons showed significantly more improvement in literacy skills than those following clustered sessions. From their data, the researchers conclude that results from laboratory experiments on the spacing effect can be applied to classroom settings in order to improve learning.

In the next section several programs in high schools and colleges will be presented in which models that concentrate the hours of instruction have been implemented and compared with regular classroom programs, in which the hours of instruction were distributed over longer periods.

3.2. Massed and distributed learning: From the lab to the classroom

Although the different educational models which involve alternative time distribution to the traditional schedule were not designed to test the spacing effect in educational settings, research on these types of programs can certainly provide empirical evidence on how time distribution of instruction hours affects the

students' acquisition of new knowledge. In Canada, and especially the US, several high schools have been offering a type of instruction which differs from the traditional schedule in being more concentrated (block scheduling). Similarly, many subjects have been taught in American colleges in an intensive way for different practical reasons. First, a review of block scheduling will be offered in this section; after which, concentrated instruction in college will be examined.

Block scheduling is a high school model which was generalized in some American high schools in the early to mid 1990s, although its origins can be traced back to the 1960s. This model was inspired by Joseph M. Carroll, who observed that academically troubled students in remedial summer school programs in the District of Columbia made gains in English and mathematics (which they studied for 4 hours a day, 5 days a week during 6 weeks) comparable to students in 2-year regular classes. Additionally, attendance was higher and students' attitude better than in regular classes. Similar results were reported in an equivalent program in New Mexico Public schools some years later (J. M. Carroll, 1994). J. M. Carroll's desire was to replicate this model used for remedial summer schools in the regular classrooms. He named his model the Copernican Plan, since, according to this author, it represented a revolution in high school education similar to Copernicus' revolutionary explanation of the movements of the planets (J. M. Carroll, 1987). Moreover, it is not simply a different time distribution which characterizes block scheduling, but there are also other methodological and practical changes involved, which are meant to promote the creation of an environment that fosters improved relationships between teachers and students and creates manageable

work conditions for both (J. M. Carroll, 1994). Those conditions include smaller groups, fewer students with whom teachers have daily contact, development of seminars, etc. Even though this schedule was adopted by some schools in the 1960s, it was not until the 1990s that the model was extended, after the National Education Commission on Time and Learning (1994) reported that there was some problem with the timing devoted to learning in the traditional educational system, claiming that learning in America was a 'prisoner of time'.

Despite the fact that other methodological changes are necessarily adopted with block scheduling, such changes could not be accomplished without a different time distribution. Several models have been proposed and implemented (Irmsher, 1996), which differ from the traditional American high school scheduling, which typically includes six to eight 40-45-minute periods a day. This traditional schedule was based on the idea that learning consists of memorization, and lectures were the best teaching methods (Cushman, 1989). J. M. Carroll (1987) claimed that, as a consequence of the different beliefs about learning adopted towards the end of the 20th century, different high school programs should be promoted. Among the different block scheduling models, the most common have been the following:

- 4x4 block: the students take four classes each semester, for which they meet for approximately 90 minutes a day every day. The former year-long courses are taken in only one semester.
- 8 block or A/B block: the students take seven or eight courses per semester; they meet on alternating days for approximately 90 minutes per session and the students take the same classes during the whole academic year.

Among the benefits of block scheduling one can highlight the fact that there is less time spent on class changes, and as a consequence there are also less problems regarding discipline, since most of such problems occur during breaks (Queen, 2000; Wallinger, 2000). Additionally, students' grades, attendance and graduation rates have often been claimed to increase on block schedules (Lewis et al., 2005; Wallinger, 2000). Since larger blocks of time are allocated for each subject, time-on-task is higher (J. M. Carroll, 1987), and a deeper investigation of the different issues dealt with in class is possible. Furthermore, different teaching interactive techniques which satisfy different learning styles can be practiced (Irmsher, 1996).

Nevertheless, there are also some negative effects of block scheduling. For example, many authors have reported that there are fewer total hours of instruction than in the traditional model (de López, 1996; Lewis et al., 2005; Wallinger, 2000), especially when two of the traditional 50-minute classes are replaced by one 90-minute class (which means 10% less of class time). Nonetheless, Queen (2000) reports that some teachers have found that the loss of instruction hours is negligible and the coverage of the content is not highly reduced with respect to the courses in the traditional schedule. Moreover, if the time taken to perform administrative tasks (such as checking attendance) before each session is considered, it becomes apparent that less time is used for such functions in block-scheduled classes (since there are fewer sessions) than in traditional. Consequently, such gains in time can compensate for the slightly reduced time allotted for each subject in block scheduling.

An important problem concerning block scheduling is the gap between courses with related content, especially languages (Queen, 2000; Wallinger, 2000). If a student is learning a foreign language in a concentrated schedule for 4 months but does not continue practicing it for the subsequent 8 or 9 months, a lot of what was learned will be forgotten, since foreign language learning is one of the areas which require long, uninterrupted study (de López, 1996; Wallinger, 2000). Other disadvantages of block scheduling include the fact that students have to make an extra effort to pay attention for a longer time span; besides, they need to devote more time to homework assignments, and keep up with the faster pace which characterizes this type of instruction (Lewis et al., 2005; Wallinger, 2000). In addition, those students who cannot attend class have a harder time to make up for the content missed. In fact, Van Mondfrans, Schott, & French (1972) claim that block scheduling requires considerable maturity on the part of the learner, as reflected in their study where only seniors were observed to make more progress in block scheduling when compared to those students in the traditional program. Furthermore, the authors mention that block scheduling can be effective for 'good' students, but not so much for 'poorer' students.

As Lewis et al. (2005) mention, there has not been much quantitative research analyzing the performance of students in block scheduling and comparing their results with those of learners following the regular high school program. Most of the reports related to this school model come from the impressions of the people involved, such as principals, teachers, students, parents or school administrators, who are usually claimed to favor block scheduling.

However, not all the opinions are positive. For instance, some teachers mention the fact that it is hard to maintain the students' attention for such long time spans. On the part of the students, the complaints mainly come from the amount of work that block scheduling entails. The gap between continuing courses is also considered a problematic aspect of this model.

The few quantitative studies which have been performed as regards block scheduling have reported mixed results, although the general picture is that there is not much difference in test scores between students in block and traditional scheduling. Van Mondfrans et al. (1972) compared the scores of students attending block scheduling and traditional programs who were in grade 9, 10, 11 and 12 through teacher-made tests and ratings of students' attitudes towards learning. Their results show that only for the seniors (grade 12 students) was block scheduling more effective than the traditional classroom program for all the subjects considered (English, Physical Science, Biology, U.S. History, Government/Economics). With respect to students' attitudes and interests, no difference was found between the two program types.

When analyzing the performance of students in French I in grade 9, Wallinger (2000) found that those learners following the traditional schedule get significantly more hours of instruction than those in block scheduling (in the two models, 4x4 and alternating day schedule). Despite this advantage that the students in the regular French classes had in terms of total number of hours of instruction, no significant differences were found in the performance of the students in the two programs in speaking, writing, reading or listening.

Nevertheless, in a more detailed analysis, Wallinger (2000) observed that the students in the 4x4 block schedule tended to cluster in the lowest quartile with respect to their scores on the reading and listening test.

In their analysis of the empirical studies which have examined the effect of block scheduling and compared it to the regular high school program, Lewis et al. (2005) reach the conclusion that, in general, block scheduling has been reported to have a minor negative effect, yet such effect is unlikely to have any practical consequences. The authors suggest that block scheduling should be considered as a part of a larger effort meant to improve high school education; however, the effects of such program are not clearly visible in the first few years of implementation. Moreover, block scheduling has to be accompanied with other different reforms in the educational system (smaller classes, different methodological strategies, more individualized instruction, etc.) in order for it to be truly effective.

The equivalent of block scheduling at universities would be the accelerated courses, which typically offer instruction in a concentrated period of time during slightly fewer hours than the traditional university classes. These intensive courses were created mainly in order to serve the needs of those adult professionals who were interested in attending university, yet, due to work schedules, could not devote long periods of time to such endeavor. These courses have been offered for a wide variety of subjects, ranging from foreign languages to economics, earth science, educational psychology or research methodology in a high number of colleges across the U.S. (Nixon, 1996). The intensive courses at university first developed during summer sessions; later, they were offered in the form of

semester intersession courses, or weekend courses (Daniel, 2000).

The results achieved by the students following intensive instruction in college have been reported to be comparable to those obtained by their peers in regular courses (Austin, Fennel, & Yeager, 1988; Caskey, 1994), or even superior (Seamon, 2004; Van Scyoc & Gleason, 1993). It is also true, though, that many studies which reported higher scores on tests at completion of intensive courses as compared to regular courses also claimed that the significant difference was not maintained over a long-term period. For example, Seamon (2004) analyzed the performance of two groups of students enrolled in a graduate educational psychology course. One group (31 students) registered in an intensive course (3 hours a day, 5 days a week for 3 weeks, over a total of 45 hours of instruction), whereas the other (35 students) followed a typical semester-length course, meeting 3 hours a day, 1 day a week, for 15 weeks (also receiving a total of 45 hours of instruction). The posttest scores were significantly superior in the case of the students in the intensive course, and such superiority was not due to the students' individual characteristics (motivation, age, or GPA), since the participants were similar in terms of those features. Nevertheless, the significant superiority on the part of the students in the intensive course was not maintained after three years had elapsed, as shown by the results obtained by a subset of participants (6 for the intensive and 9 for the semester-length course). This latter result, however, should be taken with caution owing to the size of the groups in the follow-up study.

Even if Seamon (2004) reported no differences between the students in intensive and regular courses in terms of motivation, age, etc., many researchers

have described the students in accelerated programs as being more motivated and older; moreover, they are claimed to have higher grade point averages. In addition, students in intensive courses have been reported to have a specific learning style which makes them more apt to learn in an accelerated manner (Daniel, 2000; Wlodkowski, 2003).

The teachers in concentrated courses are also expected to make a special effort in terms of class preparation and class methodology: a variety of teaching methods and activities should be used in order to keep the students engaged and create a comfortable environment (Scott, 1996). Besides, instructors are encouraged to use a different type of assessment (Scott, 1996; Watson, 1998).

When analyzing the opinions of students and teachers regarding these accelerated courses, it can be observed that most students generally favor this format, since accelerated courses are more convenient (especially because of the short time commitment), they stimulate deeper discussion due to the longer sessions; moreover, in these courses the teachers usually demonstrate creative teaching (Daniel, 2000). Scott (1995) presents a deep analysis of students' learning experiences in intensive courses as compared to regular semester-length classes. The author reports that students in intensive courses believe that such courses, when properly implemented by motivated instructors who use a variety of effective teaching methods, "can create a more focused, collegial, relaxed, motivating, concentrated, memorable and continuous learning experience" (Scott, 1995: 207). This positive feeling towards intensive courses can certainly translate into better learning outcomes. Nevertheless, intensive courses are also perceived

as more stressful than regular courses (Daniel, 2000), because of the more concentrated effort on the part of the students and, sometimes, due to the lack of time the students experience when they need to prepare for the classes. On the part of the faculty, accelerated courses are controversial, since many instructors consider such courses too demanding (Daniel, 2000). Moreover, some academics have criticized intensive instruction, as they believe it lacks the rigor and depth of semester-length courses (Wlodkowski, 2003).

Considering the reports on block scheduling in high schools and accelerated courses at universities, it can be stated that both types of programs have been controversial, both having supporters and detractors. Commonly, the students and teachers involved in the courses have been claimed to favor programs providing more concentrated hours of instruction, especially in the case of university students, because such courses demand a short time commitment. More importantly, in general, the reports on students' outcomes on these non-traditional programs have declared that the skills attained by the students in intensive courses are at least comparable to those attained by students in regular programs. Moreover, some studies have demonstrated superior students' outcomes at the end of an intensive course, though such superiority is not always maintained after some time has elapsed.

Once the implementation of intensive instruction for courses other than foreign languages has been reviewed, the next section will concentrate on the experiences with intensive instruction in the case of language courses.

3.3. Time distribution and language learning

3.3.1. The origins of the intensive programs

Although intensive programs had been considered since the first years of the 20th century (Hills, 1919), the first time that intensive language instruction was given considerable attention, as well as innumerable resources, was during World War II under the Army Specialized Training Program (ASTP) in the 1940s in the United States. Because of the immediate need the U.S. Government felt to have a high number of soldiers proficient in foreign (especially European) languages, several programs were designed in some American colleges so as to train these 'unusual' language learners with the maximum efficiency, which generally implies the minimum time. The term 'intensive' as applied to these programs usually has different connotations. First of all, intensity is related to time concentration, normally in addition to an increase in instruction hours (Pargment, 1945; Springer, 1944). Some authors consider this time concentration as the main characteristic of the intensive programs inspired by the ASTP (Kilker & Gunderson, 1972). On the other hand, other authors, who have even classified this program as the 'Intensive Method', consider that time distribution is not the main feature, but the mimicry and repetition which was promoted in such courses (Fuerst, 1944; Ittner, 1944). In this respect, Girard's words are especially significant (1943: 27):

> Everlasting practice and repetition are the most important factors. Very true. But these take time, and language learners need plenty of it whether they study for 160 school hours spread out over two years of classroom instruction, or

> whether they study for 160 hours at Fort Laramie in concentrated doses of four hours a day.

For other authors, what defined the intensive courses in the ASTP was its emphasis on aural-oral skills, as well as a determination on both teacher and student on the acquisition of speaking fluency, again irrespective of time distribution (McMullen, 1950). As a consequence, it must be born in mind that intensive language instruction in the ASTP, and the colleges which later followed this approach, differs from other types of instruction not only in its concentration of class hours, but also in terms of the specific methodology used, which formulated the principles of the audiolingual approach.

Although some authors have highlighted the diversity in terms of how foreign languages were taught under the ASTP (Paulsen, 1945), there are some common, central characteristics of this approach. With respect to the main objective, there is a clear emphasis on the acquisition of aural-oral competence, the focus being on the 'colloquial form of the language' (Springer, 1944). Second language learning was claimed to be similar to first language acquisition with respect to the fact that first one has to train the ear (listening), then the tongue (speaking), then the eye (reading), and finally the hand (writing) (Agard, 1946).

Regarding organization, the average contact hours each week would be 16 distributed over approximately 9 months. Out of those 16 hours, 5 would be devoted to the presentation of the material in groups of approximately 20 students, and 11 would be used for practice in smaller groups of around 10 students.

As for the teaching method, there was an emphasis on repetition and drill-

work. Language learning was thought to consist in the acquisition of linguistic habits, and, as stated by Ittner (1944: 180), "a habit cannot be acquired except by drill, by constant repetition; moreover, language learning is basically physiological, mechanistic –not intellectual." There is an obvious connection between language teaching in the ASTP and the main psychological trend at the time (behaviorism). The students initially memorize what the instructor presents through drills, and then apply such knowledge to natural situations. Since repetition can be quite dull, changes of activities in the practice sessions (debates, dramatization, or use of supplementary aids, such as records, radio shows, etc.) were encouraged. Grammar rules were explained inductively, and this language area, along with writing, was considered a secondary objective, hence not much time was devoted to the development of those skills.

With respect to the instructors, each student had at least three, one for the presentation session, who ideally was a bilingual speaker of English and the L2 being taught, apart from being a skilled linguist. Then, there were at least two other instructors (who had to be native speakers of the L2) for the practice sessions.

In order to participate in the ASTP, the students were selected on the basis of their language aptitude, and all of them had a clear common objective. They were encouraged to use the L2 outside the class with their peers and certainly with their instructors.

Apart from the accelerated foreign language instruction under the ASTP during World War II, there were also intensive courses at the Navy School of Military Government and Administration, which had basically the same objectives,

principles and methodology. The daily contact time was more concentrated for the Navy, with 5 hours a day, during a range of 8-18 months (Kalivoda, 1975). After World War II these intensive programs continued being offered for government and military personnel in the Army Language School, which later became the Defense Language Institute, or for the Peace Corps.

There are different opinions with respect to the efficacy of the intensive programs offered for the American soldiers in the 1940s. While there is a general agreement that such courses responded to a specific need and awakened the population about the importance of learning foreign languages, some authors have claimed that the highly positive reports on the exceptional effects of such programs gave the wrong impression about how easy second language acquisition was (Pargment, 1945). Several reports asserted that some soldiers, after learning the foreign language for a few months, were able to discuss political or legal matters using the L2 with the natives. Whereas those claims might have been exaggerated and probably unreal, some more realistic reports indeed suggested that, after a nine-month training, the students learned to understand the foreign language as spoken by the natives on a variety of topics, could speak intelligibly, read with considerable facility and write with reasonable skill (Agard et al., 1945).

Even if the majority of the intensive courses under the ASTP lasted for 9 months, some highly positive results were reported for even a three-week course (Leavitt, 1943), which was provided for some soldiers who unexpectedly arrived before the program was scheduled to start. The report claims that after such short intensive course the students had a fair working vocabulary, were able to

understand the L2 (Spanish) when spoken slowly and could express themselves with considerable confidence.

Due to the general success of the program, similar programs began to be developed for civilians in colleges. Agard et al. (1945) in their *Survey of language classes in the Army Specialized Training Program* openly encourage an adaptation of the intensive courses to American colleges, elementary and secondary schools. In fact, the approach used in the ASTP attracted many teachers, who already believed that in order to learn a language, more intensive exposure than three hours a week was necessary, as well as a practical application of the knowledge of such language in speaking activities (Zeydel, 1946).

Many colleges in the United States started to design intensive language courses for their students, with the corresponding problems in the adaptation of the approach. Especially problematic was the provision of the amount of contact hours the soldiers received, both in terms of total time devoted to language learning, and especially regarding the concentration of instruction hours (Agard et al., 1945; Padín, 1945; Santosuosso, 1957). Such problem was sometimes overcome by the implementation of semi-intensive courses, which offered between 6 and 10 hours of foreign language instruction a week, and distributed over two semesters the language instruction equivalent of two academic years. Goedsche (1946) reports on a semi-intensive German course offered at Northwestern, for which he mentioned some advantages with respect to the traditional course: the students practiced more often their speaking and listening skills, and consequently acquired more fluency. Moreover, since the groups were smaller, more personal instruction

could be provided. As drawbacks, the author mentions the fact that no literary texts were read (reading literature in the L2 was one of the main activities under the grammar-translation method, which was the most popular approach before the implementation of the audiolingual method). Additionally, some students complained about the increased number of classroom hours a week as compared to regular courses (however, this fact is compensated by fewer hours of homework). Oswald et al. (1950) also report on a successful German semi-intensive course at UCLA, where the students outperformed their peers in regular courses, especially in reading and vocabulary, yet their grammar skills were not so clearly superior.

Although according to a few authors semi-intensive courses are more effective than intensive courses, since the former are less demanding and not as concentrated (Nordsieck, 1946), there seems to be a general agreement that one of the positive features of intensive courses is that the students' learning of a foreign language is not disrupted by the learning of other subjects. When evaluating the success of their intensive program, Kilker & Gunderson (1972) claim that the main aspect of such success was not the method or the materials used, but that the students had a "large block of time unencumbered by other course work" (p. 850). Cowan & Graves (1944) suggest that there should be at least 10 hours a week of contact to obtain improved results in language courses.

In order to make arrangements for time distribution easier in the case of intensive programs, some institutions offered these courses during the summer. Metcalf (1946) reported on the higher achievement of the students registered in the intensive summer program (12 hours a week), as compared to traditional courses

(4 hours a week); furthermore, these learners' results were even higher than those obtained by the students in the intensive course offered the previous summer, which provided 9 hours a week of instruction. Similar success is claimed by Moore (1950) in an even more intensive summer program (20 hours a week over a total of 120 hours). Both authors also mention some disadvantages of intensive courses, namely the fast pace (which for some students is difficult to follow), and the fact that the students who register in such classes need to be motivated and committed to work hard. Additionally, the instructors need to have an inspiring personality and use a variety of techniques to break up the monotony that so many continuous hours of instruction could cause.

While the general feeling about the intensive programs for military personnel was one of success (Agard et al. 1945; Shueler, 1944), the opinion about the positive value of the adaptation of such programs in college was controversial. Intensive courses implemented in different American colleges and universities were usually claimed to produce more positive results with respect to language learning than the traditional 'dribble method', which provided a maximum of 3 hours a week of contact with the foreign language (Cowan & Graves, 1944; Springer, 1944). Nevertheless, there was some criticism about the appropriateness of such courses in an academic environment. First, the high amount of contact hours in a limited time period was considered problematic. Apart from reporting on the difficulty of adapting these programs in terms of hours of instruction to elementary, secondary school or college, Padín (1945: 63) presents his picturesque opinion about intensity in language learning:

> Es verdad que el aprendizaje de una lengua, lo mismo que el buen vino, exige, además de aplicación, tiempo, porque sin el tiempo no se va a ninguna parte ni se está en ninguna parte (…) el método intensivo no ha creado píldoras de vitamina lingüística que uno puede tomar por la mañana, con el desayuno, y salir hablando chino o japonés a la caída del sol.[15]

Pargment (1945) suggests that, under normal circumstances in which there is no pressure to learn a language in a limited period of time, a less intensive approach should be preferred in language learning, since having shorter intervals between sessions hinders the assimilation and retention of what is learned (still, he concedes that an advantage of intensive exposure is that it reduces distraction and forgetting between sessions).

Other authors, more academic in their orientation and probably aligning to what were before that time the most commonly used teaching approaches (grammar-translation and the reading approach), criticized the teaching methodology used in the intensive army program, since it simply promoted mechanistic memorization, which was believed to produce only the limited results for which the ASTP was designed. Conversely, language learning in college should be a more intellectual process; the students should be encouraged to reflect about the language, and read foreign literature (Oswald et al., 1950; Rogers, 1945). Moreover, there should be higher standards of accuracy than what was expected

[15] It is true that language learning, the same as a good wine, requires, apart from effort, time, because without time you cannot go anywhere and you cannot be anywhere (…) the intensive method has not created linguistic vitamin pills that one can take in the morning, with breakfast, and end up speaking Chinese or Japanese by sunset.

under the ASTP (Pargment, 1945). Brann (1944) presents a more balanced position, claiming that the repetition method and the long exposure to the target language characteristic of intensive programs is positive for the beginning stages of learning; however, when the student's knowledge is more advanced, a more intellectual analysis of the language should be encouraged.

Wilson (1965) points out that in order for the intensive program to be successful at universities, certain criteria should be met for the instructors, materials, and block of contact time, to make sure that the students not only get concentrated material, but also the opportunity to develop new linguistic habits. This author criticizes the amount of intensive programs which were developed in different colleges not taking into account the implementation issues he emphasizes.

Most of the literature on the intensive programs developed during and after World War II tends to be quite impressionistic when reporting the results of the model, and most authors merely present their own beliefs about language learning. There are few studies which certainly present results of tests and comparisons between intensive and non-intensive programs at university. Such studies (Moore, 1950; Oswald et al., 1950), despite not being highly precise about their methodology in data collection and analyses, tend to show higher scores on tests for students in more intensive programs, especially in vocabulary and speaking skills.

In sum, the studies presented above which have discussed the impact of the intensive language programs developed under the ASTP on foreign language

proficiency tend to suggest that the learners can make linguistic gains in a short period of time. The positive effect of the intensive model is claimed to be higher for students at the beginning stages of language acquisition than at the advanced (Schueler, 1944; Springer, 1944), despite the fact that the opposite claim was also made by Pargment (1945) who considered advanced learners to be better prepared for the intensive experience[16]. The gains reported refer mainly to aural-oral skills, which were the ones that received the most attention in the classes following the ASTP.

The positive results obtained by the soldiers in the ASTP or by the students in the colleges where such approach was implemented could be due to time concentration, increase in instruction hours, the methodology used in the class (audiolingual approach), aptitude, or motivation on the part of the students. Indeed, in many cases students were selected on the basis of their motivation or aptitude in order to participate in the program. Additionally, Menut (1953) believes that screening students on the basis of these characteristics should be done if positive results are expected. Despite the fact that the results of the students following an approach similar to the ASTP in colleges are controversial, there seem to be many claims that such approach enhances students' language skills as compared to other approaches (Moore, 1950).

In the next section other versions of the intensive model during the second half of the 20th century will be presented. As will be observed, most of the

[16] It must be born in mind that authors do not provide a clear definition of what an 'advanced' learner is.

programs which will be reviewed concern adult SLA.

3.3.2. Intensive language teaching in the 1960s-1980s

During the 1960s, 1970s and 1980s it is again in North America where we find research being done regarding the effect of time distribution on L2 learning. What was attractive to the American public about the intensive courses was the time concentration, irrespective of the methodology which was followed (in fact, not all the methodological principles and objectives formulated for the ASTP were maintained during these decades). It was the appeal of acquiring a certain competence in a foreign language in a minimum amount of time (which is claimed to be in accordance with the 'American spirit' of impatience and search for immediate results) that attracted students to intensive language programs (Benseler & Schulz, 1979). As Schneider (1977: 27) reports, "Today's students do not want to invest six years in becoming proficient in a foreign language; they want to learn it today and speak it tomorrow". Despite some difficulties in the implementation of highly intensive programs due to time restrictions, Kalivoda (1975) encourages the creation of such courses at colleges and high schools, in order to offer the students the possibility of choosing the program type that best fits their learning styles.

In their analyses of intensive foreign language instruction, Benseler & Schulz (1979) present the different types of programs which can be considered 'intensive', as well as a good description of the main components that such

programs need to have. Among the different modalities of intensive programs, these authors mention intensive courses for specific skills (listening, reading, etc.); courses for language teachers; language weekends, retreats, camps, etc.; intensive programs for special purposes (business, medicine, etc.) for service oriented groups (Peace Corps) etc.; intensive study at high school; intensive preparation in foreign languages previous to college; immersion and study abroad courses; and finally, immersion experiences provided by residential facilities (special language houses or dormitories).

The most common types of intensive programs in college are summer courses, usually involving housing facilities for the students where they can (and in many cases are required to) continue practicing the foreign language outside the class (Currall & Kirk, 1986; Keilstrup, 1981; McKee, 1983; Schneider, 1977). Other common types of college intensive language programs include semester-length intensive courses (Frank, 1972), or semi-intensive courses (with fewer than 10 hours a week) (Cipolla, 1982). It is emphasized, though, that it is highly important that the students do not take any other subject during the time they are attending the intensive language classes (Kalivoda, 1975; McKee, 1983; Schulz, 1979), since the concentration on the language study would not be as exclusive and intensive as desired. Students are aware of the importance of concentrating on language study only, as reflected in the opinion of a learner in a summer intensive course: "It was better to concentrate on one subject and not be distracted by other courses" (McKee, 1983: 11).

Concerning the levels being offered through intensive instruction, most

programs include beginning or intermediate levels (Benseler & Schulz, 1979; Byrd, 1980; Frank, 1972; Marti, 1972; McKee, 1983), while very few reports exist on intensive advanced language programs (Keilstrup, 1981). Such distribution may reflect the implicit assumption that intensity has more positive effects at more initial stages in the language acquisition process, or it may simply be due to greater demand of foreign language instruction at beginning stages. Paquette (1973) even suggests that beginning and intermediate foreign language instruction should be offered exclusively in total immersion programs. Gardner, Smythe, & Brunet (1977) in their analysis of the progress in oral skills by beginning, intermediate and advanced learners of French in an intensive program found that the higher the students' initial level was the less progress was experienced at the end of the course in terms of oral fluency.

What intensive courses have in common is an increase in contact hours with respect to the otherwise normal classroom period, which is the main characteristic of intensive instruction according to Benseler & Schulz (1979). Such programs do not need to follow a specific linguistic theory, methodology or use specific teaching materials; nevertheless, a lot of them continued using some version of the audiolingual approach, emphasizing the importance of drills and repetitions (Byrd, 1980; Frank, 1972), or the practice of the aural-oral skills (Marti, 1972), although in many cases all the different language skills were practiced (Frank, 1972; McKee, 1983), and more importance was given to grammar, writing and reading than under the audiolingual approach.

Benseler & Schulz (1979: 9-10) mention the following aspects as necessary

for the successful implementation of an intensive program:

- Extended daily exposure to the language (from 2 to 8 hours a day)
- Larger number of classroom contact hours (increases up to 75% in instructional time)
- Usually small classes, 10 students maximum (in big classes practice is promoted in small groups)
- Focus on oral/aural skills, spoken language, but also some attention to the grammar necessary for clear communication in L2
- Instruction is implemented in the foreign language
- Frequent use of the language laboratory for practice
- Extra-curricular activities in the target language
- Staff willing to devote more time and energy than what is required for regular courses

Apart from the 'time advantage' that intensive programs offer, some authors considered that these programs were highly positive due to the fact that they resembled natural language learning situations, because of the immersion that takes place in the foreign language (Benseler & Schulz, 1979; Marti, 1972). Additionally, the students develop a camaraderie in these programs due to the concentrated time they spend together, which is a positive factor that probably facilitates the learning process (Benseler & Schulz, 1979; Moreland, 1980)[17]. Most importantly, intensive instruction was also claimed to provide learners with

[17] Hinger (2006) also found a similar positive effect of intensive instruction in students' relations, which had an impact on their motivation to learn the foreign language.

superior language proficiency than the traditional courses, which offered a maximum of four hours a week over approximately four semesters (Deveny & Bookout, 1976). Furthermore, after completion of an intensive program students were reported to be more motivated to pursue language studies or activities for which the knowledge of a foreign language is necessary or desirable, such as spending periods of time abroad or studying the foreign language literature or culture. Gardner et al. (1977) observed that after an intensive French course the students were more motivated to learn the language and felt more at ease with it.

During the years comprised between 1960 and 1990 there is a dearth of empirical studies presenting clear evidence for the superior language proficiency of the students in intensive programs as compared to those following the traditional schedule. This fact was already underlined by McKee (1983), who emphasized that such studies were necessary in order to present some evidence for the claim held by many authors that intensive programs were beneficial for the students who were learning a foreign language.

In her study, McKee (1983) compares two groups, one (with 11 students) receiving intensive French lessons for six weeks during the summer (approximately 25 contact hours a week, over a total of 120 hours), and another following the traditional schedule (20 students who had approximately 4 contact hours a week during two quarters, over a total of approximately 120 hours). The results of the study show that the students in the intensive program outperformed their peers in the traditional setting in free composition, listening, reading and writing, yet the difference between the two groups was only significant in the free

composition task (which she analyzed in terms of 'idea units'[18]). Acknowledging the limitations of her study (an important one being the low number of participants[19]), McKee (1983) recommends the promotion of intensive programs and presents as a positive outcome of such programs the fact that the students enrolled might be more eager to use the language (at least in writing), as seen in their significantly superior performance in the free composition task.

There are other earlier studies which report on results of students in intensive programs with respect to those registered in traditional classes. Williamsen (1968) compared 19 students in a summer intensive program who received 8 weeks of Spanish instruction (approximately 25 hours a week) with students (195) enrolled in the traditional 4-semester Spanish course (approximately 4 hours a week during two academic years). Even though the results of the MLAT[20], the questionnaire about foreign language experience and performance on the SATs[21] (verbal and mathematical) were slightly more favorable to the traditional group, the MLA[22] Spanish Level L test (which was typically given at the end of the second semester of Spanish) showed similar scores for the students who had been enrolled in the intensive summer program and those who had followed two semesters of Spanish in the traditional program.

[18] McKee (1983: 17) defines an idea unit as "a main idea or any object (direct or indirect) or any accompanying modifier (clausal or nonclausal)."

[19] This is one of the limitations she mentions, yet I consider another limitation the lack of pretest.

[20] Modern Language Aptitude Test developed by Carroll and Sapon during the 1950s in order to measure students' aptitude for learning foreign languages (Carroll & Sapon, 1955).

[21] SAT stands for Scholastic Aptitude Test, and it is the entrance test for U.S. colleges.

[22] Modern Language Assessment: test used to measure proficiency in foreign languages which includes three parts (listening, reading and writing).

The best students in the intensive group (12, or 60%) were chosen to take the MLA Spanish Level M test (typically given at the end of the fourth semester of traditional Spanish instruction). The results obtained by such group were similar to or higher than (significantly higher in writing) the scores of the students who had finished their fourth semester of traditional Spanish lessons.

Williamsen (1968) does not specify the amount of contact hours the students in the traditional program had before taking the L or the M exam, as compared to the intensive group. Probably, what Williamsen (1968) wanted to emphasize was the fact that, just by investing eight weeks in the summer, many students could achieve a degree of proficiency in Spanish comparable to the proficiency level attained in four semesters (the equivalent of two academic years).

A similar comparison was made by Frank (1972) with respect to the acquisition of German as an L2. He reported that the scores of the students in the intensive program (which in one semester offered 300 contact hours, approximately 20 hours a week, half of which were in class and the other half in the language laboratory) were 35% higher than the scores of the students registered in the traditional German four-semester course in a test, developed by his department, which evaluated reading, grammar and listening. Moreover, the students in the intensive program scored 25% higher than their peers in the regular program in a test from the Educational Testing Service of Princeton, which measured reading comprehension and grammar. Apart from obtaining better scores on language tests, the students in the intensive course demonstrated higher

satisfaction (despite the hard work) than the students in the regular classes, and many of them pursued other German courses. Frank (1972: 32) concludes: "Maximum exposure in the shortest possible time can provide the answer for undergraduates who question the validity of a four semester foreign language requirement that fails to give them a functional control of the language."

In addition, there are other studies which, even without providing a comparison group, report on the achievement of students in different intensive courses. For instance, Marti (1972) provides some information regarding an intensive French program, which lasted for one year (instead of the two in which the regular program was implemented). In such program, a total of 10 classroom hours were provided each week. Moreover, the students were required to work in the lab for 5 hours. Five more hours were spent practicing oral skills with a French-speaking assistant. Additionally, the students were expected to do many hours of homework; consequently, the total amount of contact hours was around 30 a week. Marti (1972) found that the scores of the students on the MLA French test were outstanding, with more than 80% of the 64 students obtaining a score of 81% or higher.

Similar encouraging results were reported by Schneider (1977). He analyzed the performance of students who had attended a German intensive course for seven weeks during the summer (approximately 20 contact hours a week, which was the equivalent of two college semesters). The average results of the MLA Test in German and the CLEP (College Level Examination) test in reading, listening and grammar were usually above 90%.

Keilstrup (1981) presents the results obtained by a group of students in an intensive advanced course, which is one of the few studies including a level other than beginner or intermediate. He claims that the proficiency level of students after completing an intensive German course was 9% higher than that of the average traditional course (no information is provided as to the type of test, the scores obtained by either group, or when the test was performed). Furthermore, the majority of the students (79%) were reported to have obtained 80% or higher grades on their tests (again, there is no specification regarding the type of tests).

The advantages which have been presented in favor of intensive programs (limited time commitment, higher proficiency levels attained, increased motivation on the part of the students) do not come effortlessly. Both the instructor and the students need to make a commitment to work harder than under the traditional schedule, and be enthusiastic about teaching and learning respectively (Schulz, 1979).

There is a type of intensive program (called the Rassian Method or the Dartmouth Intensive Language Model) which highly emphasizes the role of the instructor as an enthusiastic and uninhibited person. This method was developed by John Rassias as an adaptation in his college (Dartmouth) of a program he developed to teach volunteers at the Peace Corps in 1965. This model provided an increased amount of hours with respect to the traditional schedule (approximately a total of 20-30 contact hours a week). There were 10 hours a week of instruction, which were distributed between master classes with experienced teachers (5 hours a week), practice, and drill classes with apprentice teachers. Then, two hours were

spent at the language laboratory, where the students did individual practice. Another inherent characteristic of the method was the practice of the audiolingual approach, yet a 'revitalized' version of it (Luplow, 1982), since classes included dramatic performances and the instructor was expected to recreate real-life situations in a theatrical way; that is why, for the correct implementation of this method, the teacher had to be enthusiastic and uninhibited. The Dartmouth Intensive Language Model was implemented (usually successfully) at different universities (Cipolla, 1982; Luplow, 1982; Winston & Boots, 1987), although sometimes the model did not have such a positive effect due to the special circumstances of the college where it was implemented (Petadella, 1988).

As was mentioned above, the students were also expected to have certain characteristics in order to succeed in an intensive program. First, special aptitudes and learning skills are highly desirable; among which language aptitude is especially important, which was measured for the students in many intensive programs (Williamsen, 1968). According to Carroll (1974), another important skill is the ability to learn fast; intensive programs are not recommended for students who learn at a slower pace.

Apart from the scores on aptitude tests (especially the MLAT), other academic scores were considered as requirements for participation in some intensive programs, such as SAT's (Williamsen, 1968), ACT[23]'s (Marti, 1972), or the students' grade point average of their academic record in college (GPA) (Currall & Kirk, 1986). Curral & Kirk (1986) did some research in order to investigate which

[23] College entrance test designed to assess students' general educational development.

factors were most likely to predict success in an intensive French summer course. After correlating the grades obtained at the end of the course with other variables, such as sex, age, previous knowledge of the language, previous music training, reasons for studying French, interviewer's rating of the applicant's aptitude for success, etc., Currall & Kirk (1986) concluded that a multiple regression using three predictors (GPA, interviewer's rating of the applicant's aptitude for success and previous language knowledge) can effectively predict course grades in the intensive program.

Despite the fact that many studies highlighted the importance of certain aptitudes on the student's part, other researchers claim that lack of aptitude can be compensated with motivation. Benseler & Schulz (1979) and Schulz (1979) argued that motivation is a requirement for students to do well in intensive programs. McKee (1983) found that the students who had a more integrative motivation (as opposed to instrumental) were more successful than those who were enrolled in the French intensive course as a requirement. According to Kalivoda (1975), students should not be excluded or discouraged from taking an intensive course on the sole basis of their aptitude, since motivation has been often shown to compensate for lack of aptitude; moreover, motivation has sometimes been claimed to be the most important factor for success in an intensive language program.

In general, it can be said that most of the studies concerned with intensive foreign language programs from the 1960s until the 1980s were mostly descriptive (Hirsh, 1982; Jurasek, 1982; Sacks, 1967; Tamarkin, 1988; Urbanski, 1982), and even

those which are empirical have some methodological flaws (no pretests, low number of participants, lack of detailed description of instruments used to collect data and data analysis, no comparison groups, etc.). Despite the lack of studies which present clear advantages for students in intensive programs as compared to their peers in regular classes in a methodologically sound way, the general claim or belief which was predominant in the literature at the time was that learning a foreign language through a concentrated exposure in a limited time period was at least as effective, if not more, than following traditional language courses. Moreover, after some intensive instruction the students were reported to be highly motivated to pursue more advanced language studies. An additional advantage was the short period of time required for the acquisition of language skills as opposed to the long process necessary in traditional courses (usually involving two academic years). Although most reports highlighted the benefits of intensive instruction, some scholars have claimed that certain characteristics on the part of the teacher (motivation, enthusiasm, commitment to the program), or the student (aptitude, motivation, willingness to work hard) were highly desirable for the success of an intensive program, as well as the disposal of a certain amount of concentrated time that should be exclusively devoted to language study. An important issue that did not receive much attention was the long-term effect of intensive instruction. In general, studies did not investigate retention of the knowledge acquired in intensive language courses. It must be reminded also that all these intensive programs reviewed included adults as students, mainly in college, or in their last year of high school before going to college.

In the next section other types of intensive language programs, this time involving mainly primary school students following alternative language teaching models which are quite widespread across Canada. Moreover, since the community and the Government have been interested in the development of these programs, they have received many resources and language researchers have been asked to act as evaluators and program coordinators. Due to these favorable circumstances, data can be collected from a high number of participants, and the results obtained by these researchers, thus, have a higher reliability and are more easily generalizable than the results which have been presented in the previous sections.

3.3.3. Intensive language programs in Canadian schools

Whereas the literature reviewed in sections 3.3.1 and 3.3.2 is related to intensive programs for adults, the Canadian experiments which will be discussed in this section refer to school learners. Canada, as a bilingual country with two official (and prestigious) languages (French, a minority language, and English the majority language) has developed many programs which enrich and/or increase second language instruction. Traditional English as a second language (ESL), or French as a second language (FSL) programs, have not shown to be efficient in terms of producing highly competent individuals in these languages (Netten & Germain, 2004a; Spada and Lightbown, 1989). Since such L2 competence is decidedly promoted in this country, other alternatives have been sought. First, the

alternatives for French as a second language will be explored and then for English as a second language. Although different models that increase and/or concentrate L2 instruction will be reviewed, the main focus will be on intensive language courses.

3.3.3.1. French immersion programs and intensive French

The first immersion class in the public sector started in 1965 in St. Lambert, a suburb in Montreal, mainly because of the concerns some parents had about their children's poor French language skills after years of French instruction (Lambert & Tucker, 1972). In a short time, immersion classes were held in other areas in Canada, mainly in the provinces of Quebec and Ontario. Three different types of French immersion programs were designed: early total immersion, early partial immersion, and late immersion (Swain, 1985). The *early total immersion* is a program in which the students receive content instruction in the target language (French, in this case) as early as Kindergarten or grade 1, with as much as an average of 90% of the time in this language. After grade 6, the classes in French are reduced up to a 50% of the school time in the language. The *early partial immersion* offers the students the possibility to have content instruction in the target language from the early grades as well, yet the exposure to French is more or less 50% of the time throughout primary and secondary school. Students participating in this program typically have from 6,000 to 7,000 contact hours of French. In the *late immersion* programs, the learners start to receive content instruction in French

when they are in grade 6 or 7 until they finish secondary school, and the content classes in this language constitute more or less 50% of the total hours of class, providing typically around 3,500 hours of French input.

The evaluations of these programs have always been highly positive, since the students who were enrolled in them learned French at higher levels than those in regular French as a second language programs, with no detriment to their native language (English). Apart from this, the students showed higher cognitive abilities and a mastery of the content in the second language that was similar to those students who were instructed in their native language (Swain & Barik, 1976). Nonetheless, not all the outcomes of immersion programs have always been positive. Since the teaching approach in these programs has been to promote the implicit acquisition of the second language through content instruction or meaningful interaction, focus on grammar has been discouraged. The lack of grammar instruction may have led the students to continue making syntactic and morphologic errors even after they finished secondary school. Several researchers have undelined this fact (Genesee, 1987; Swain & Lapkin, 1986), arguing that it is necessary to devote some time to the teaching of grammar in these programs, if the learners are to overcome those syntactic and morphologic difficulties (Genesee, 1994; Swain, 1998). As Swain (1998) suggests, subject matter teaching is not always good language teaching.

Despite the problems with accuracy that students in French immersion programs have been claimed to have, this type of program has been demonstrated to be far superior to traditional FSL classes. Immersion students attain higher

levels of proficiency in French. Additionally, the students receiving this type of instruction are reported to communicate with ease in French (Lapkin, Hart, & Swain, 1991), whereas in FSL classes learners are often unable to use the L2 at the end of secondary school.

As early as 1980 there is a study (Billy, 1980) which shows that another type of enriched French language program can have even superior effects than French immersion. This program, intensive French, was developed in part due to the observation that the immigrants who were enrolling in Quebec schools learned French through intensive instruction during five months in a highly efficient way, up to the point that they demonstrated a better command of the language than English-speaking children after spending five or six years in French immersion classes.

In 1975 the Mille-Îlles school district in Montreal decided to offer a similar type of intensive instruction to English-speaking Canadians. Two experimental groups were formed (grade 1 and grade 6), which provided concentrated and intensive instruction in French in a course that increased dramatically the hours of instruction with respect to the regular traditional classes, but just focused on the language and not on the teaching of subject matter, as the French immersion classes. After analyzing the French language skills of 6-7 year-old students in grade 1 in French immersion and intensive French, Billy (1980) reports that, whereas both were comparable in terms of comprehension, the latter showed more syntactic complexity in oral production. In her comparison of 11-year old students (grade 6), equivalent results are found, namely that those students in intensive French were

producing more complex language than their peers in immersion classes, despite the fact that the latter had devoted more than 10 times as many hours to the study of French than the former. Moreover, similar types of mistakes are observed in the immersion children when they are 6 years old and 5 years later. Examining the effect of intensive French in the two groups, Billy (1980) claims that, despite such program being beneficial for all students, more profit could be obtained by the older learners than by the younger ones. Notwithstanding the superiority she claims for intensive French with respect to French immersion, Billy (1980) suggests that any of these program types is highly superior to traditional French classes, in which instruction takes place over a long period of time through small doses: "Dans les deux cas, soit par l'immersion telle qu'elle est connue, soit par des programmes intensifs, les enfants apprenaient mieux que dans les programmes à la demie-heure ou aux quarante minutes répartis sur des périodes de plusieurs années."[24] (p. 424)

Despite the claims presented by Billy (1980), intensive French did not experience much acceptance in the 1980s and 1990s due to the fact that French immersion education was considered by the majority of the population as the best alternative for Anglophone children to acquire communicative competence in French. The situation is changing at present and intensive French programs are gaining increasing support (Netten & Germain, 2004a). Although different programs had been experimented in different parts of Canada since 1975, none

[24] In the two cases, whether in immersion or in intensive programs, the children learn better than in the 30 or 40-minute programs spread over many years.

was widely accepted and extended until the end of the 1990s.

Some of the earlier experiments in intensive French include *le bain linguistique* in the Ottawa School Board during the years 1993-1996, and the block scheduling in the Carleton Board of Education in 1993-1994 (see Netten & Germain, 2004b).

Le bain linguistique offers approximately 450 hours of French instruction in one year, as opposed to the 120 hours offered in the traditional, or core French program (MacFarlane, Peters, & Wesche, 2004). Results from different research projects show that the students who were enrolled in this intensive program progressed more in their French skills than those students in regular French classes (Peters, 2000; Wesche, MacFarlane, & Peters, 1994); furthermore, those students in *le bain linguistique* program showed more self-confidence in French and also more positive attitudes towards the learning of this language (MacFarlane et al., 2004; Peters, 2000).

Another attempt at enriching core French was the block scheduling undertaken by the Carleton Board of Education. Contrarily to other intensive programs reviewed in this chapter, this model does not increase the amount of hours devoted to the second language with respect to the traditional core French model; rather, it distributes the same amount of hours differently. Two models of block scheduling were proposed: one in which the students had half days of instruction in French over a 10-week period (the half-day model), and another one which offered 80 minutes of French a day over 5 months. Compared to core French, which provides 40 minutes of instruction a day over the whole school year,

the other two models were more intensive in the sense that instruction was more concentrated than in the traditional model. Lapkin, Hart, & Harley (1998) show that the students in half-day classes outperformed the students in the regular French program, especially in reading and writing. Those students in the 80-minute program were also superior to the students in the traditional model; however, the differences in performance between the two groups were not statistically significant. These results show that the more concentrated the exposure to the second language is the more linguistic gains are obtained by the students.

After these early attempts at introducing intensive French, which were not highly successful in terms of their continuity, towards the end of the 1990s new groups started receiving intensive instruction in this language (partly following the success of intensive English) and from that time onwards, more and more schools are adopting this model in different provinces in Canada, where intensive French has become quite popular in the last few years. Netten & Germain (2004a) report on the early implementation of intensive French in Newfoundland and Labrador starting in 1998 through 2001. The researchers suggest that the outcomes of the program were quite successful and thus recommend the adoption of the model in other provinces or school boards. In fact, after 2002, intensive French has expanded to other provinces, starting with New Brunswick and Saskatchewan (2002-2003), Nova Scotia and Alberta (2003-2004), British Columbia (2004-2005), and Manitoba (2005-2006) (Netten & Germain, 2005).

Intensive French is clearly different from both, French immersion and core French classes (regular FSL courses) in a variety of ways. As already suggested,

intensive French differs from French immersion in that no subject matter is taught in the former. Moreover, whereas immersion students remain in the program for several years, intensive French is only offered for five months. Intensive French in turn is highly different from core French not only in language exposure, which is longer and more concentrated, but, according to Netten & Germain (2000), it also represents a reorientation of the curriculum, "so that it provides a rich second language experience and uses an interactive pedagogy" (p. 17). Such aim is achieved through both, an emphasis on the use of French for communicative purposes and on the implicit acquisition of the language, as well as by a transdisciplinary approach, in which issues related to content subjects are incorporated into the French class.

With respect to the first two characteristics, related to time and time distribution, intensive French provides students in grade 6 with approximately 350 hours of French in five months (usually the first five months of the school year)[25], as opposed to core French classes, which normally provide a total of 90 hours in grade 6, distributed in periods of 30-40 minutes, clearly different from the blocks of 3-4 hours a day in intensive French. In order to be able to provide students with this high number of hours of French instruction there has to be a redistribution of the rest of the subjects in the curriculum. The subject areas which are usually compressed are English language arts, science, social studies, health and religion. Another fundamental characteristic of the new approach to the teaching of French

[25] During the remaining five months the students continue their French instruction in their core French classes.

in intensive French programs is the emphasis on communicative skills and interaction, through which an implicit learning of the foreign language is promoted (Netten & Germain, 2004b).

The results of the investigations of learners' proficiency in French after attending intensive classes have been successful in terms of both oral and written production. Germain, Netten, & Séguin (2004) found that the written production from the students in intensive French programs was comparable to the written production by francophones from Quebec at grades 3 and 4. Moreover, their results suggest that some balance between accuracy and fluency can be achieved (as opposed to French immersion students who are shown to be highly fluent yet not so accurate in their L2 production). Germain, Netten, & Movassant (2004) reported similar positive results in oral production. The students who followed intensive French courses during the years 1998-2001 that they investigated achieved a score of 3.7 on a 5-point scale in an instrument developed by the Ministry of Education from Newfoundland and Labrador in order to evaluate students' proficiency in French up to grade 12. On average, the students investigated by Germain, Netten, & Movassant (2004) obtain a score quite close to 4, which stands for "Très grande spontanéité. Possibilité d'engager et de maintenir une conversation générale. Flexibilité langagière dans des situations qui dépassent les besoins de base immédiats. Utilisation créatrice de la langue."[26] Consequently, the students' oral skills were highly superior to the skills that, typically, students in

[26] Great spontaneity. Possibility of participating in and keeping a general conversation. Language flexibility in situations which go beyond immediate needs. Creative use of language.

grade 6 core French demonstrate. Furthermore, grade 6 students in intensive French are reported to achieve an oral proficiency in French comparable to grade 9-10 or sometimes even grade 11 students in traditional French courses.

Summarizing, the two main types of intensive instruction in French provided in Canada for primary school students (immersion and intensive French) consist of both an increase in the hours of instruction (which is highly significant in the case of immersion) and also a modification of the typical curriculum, which in immersion represents a clear departure from traditional language teaching, since language is learned through content subjects. Intensive French is also based on some premises that differentiate it from regular core French classes, in the sense that a more communicative approach is designed (which highlights the implicit acquisition of the language through meaningful situations). In addition to immersion and intensive French, another way to provide intensive language instruction which was presented in this section is block scheduling. It is only in this context in which the positive results obtained by the students provide evidence for the fact that solely a concentration of instruction time (without an increase in the hours of instruction or a change in the curriculum) enhances the students' L2 learning (Lapkin et al., 1998).

3.3.3.2. Intensive English

Experiments with some kind of intensified English instruction started in Quebec in the late 1970s (more specifically in 1976) in Mille-Îlles, Montreal (where

the first experiment using French intensive instruction had taken place the year before), and Greenfield Park, also in Montreal (Germain, Lightbown, Netten, & Spada, 2004). Such programs began due to similar types of concerns about the children's second language competence, although this time those concerns were coming from the French-speaking community, which was not highly satisfied with the francophone students' command of the English language. These innovative programs were also spurred by the fact that immersion in English is not allowed in French schools. The Charter of the French Language legislation specifies that all the students in Quebec will receive their primary and secondary school education in French (Lightbown & Spada, 1994). Until the mid 1980s few intensive ESL programs existed in Quebec, yet since 1985 there has been a rapid increase in the number of such programs (Spada & Lightbown, 1989), up to the point that in 2001 the Quebec Ministry of Education encouraged the use of intensive English across school boards (Germain, Lightbown, Netten, & Spada, 2004).

There are several models of intensive English instruction. In some schools the students have four months of intensive ESL and complete the rest of their curriculum in French in the remaining six months. Another alternative is offering half days of English instruction throughout the academic year. Finally, the most popular model provides intensive ESL in five months (at the beginning or at the end of the academic year), whereas the remaining five are devoted to the regular curriculum.

In this latter model (where most research has been done), English is taught at grade 5 or grade 6 during all school hours (except for art, physical education,

music and religion) for a total of 350-400 hours (Spada & Lightbown, 1989). Apart from the increase in instruction time, some schools promote the use of English outside the class. Lightbown and Spada (1997) report on the highly positive results of the intensive ESL program in a school where only grade 6 was taught and all the grade 6 students did intensive ESL, and because of these special circumstances it was possible for the students to use English outside the class throughout the school year. The authors demonstrate how the students significantly improved their listening comprehension and vocabulary in English during the semester in which they were not attending English lessons, because of the 'ambient' English, which promoted both the comprehension and use of English outside the class, while providing a motivating environment in which to practice the language.

The intensive ESL program stands in clear contrast to the regular ESL program, in which the majority of school age Canadians still receive their instruction, which in Quebec is typically 1 or 2 hours a week in elementary school, starting now in grade 1 (as opposed to grade 3 which was the typical starting age until 2006) with a total of 35-70 hours per year, depending on the school. In secondary school there is an average of 2.5 hours a week of English instruction from grades 7-11.

As reported by Lightbown & Spada (1994), in these intensive English classes the focus is on the language itself, and the students do not receive content instruction (as in immersion programs). The approach used to teach the language is communicative, with an emphasis on oral-aural skills through meaning-based activities. As was the case for immersion programs, the implicit acquisition of the

L2 is promoted, thus explicit explanations of language forms are highly discouraged. This aspect of the teaching methodology has sometimes been criticized by some researchers who suggest that occasional focus on form is beneficial for the students to improve their accuracy (Lightbown & Spada, 1994).

Results from several studies highlight the positive outcome of intensive English instruction. Spada & Lightbown (1989) analyzed the performance of 33 groups of intensive English across eight school boards in grades 5 and 6 and compared them with grade 5 and 6 students in regular English classes (who were the same age as the experimental group but had received fewer hours of instruction) and also with grade 9 students, who were older but had been exposed to approximately the same contact hours[27]. The comparison between regular and intensive grade 5 and 6 ESL programs demonstrated that the latter were significantly superior to the former in all the tests used (Baldwin-Cartier *Test de classement*, which is a test used for streaming students into beginning, intermediate, advanced ESL classes; a listening comprehension test developed by the MEQ (Ministry of Education of Quebec); and a picture card game which aimed at analyzing oral proficiency in terms of fluency and accuracy). Intensive grade 5 and 6 students were also superior to their grade 9 counterparts in the MEQ listening comprehension test. Moreover, after the intensive course, the students reported significantly more outside-of-class contact hours with English than their grade 5

[27] Spada & Lightbown (1989) mention that the comparison between grade 5-6 students and grade 9 is problematic, since they differ not only in age but also in terms of the methodology under which they learned English (the former more communicative and the latter a modified audiolingual approach).

and 6 peers in traditional ESL classes, as well as a more positive attitude towards English and the value of learning it.

Lightbown & Spada (1994) present similar results; additionally, they analyze the long-term effects of the intensive program with 60 grade 11 students (half of which had taken intensive ESL in grade 5 or 6) through a questionnaire, an interview and two communicative tasks. The students who had attended an intensive ESL class reported significantly more contact hours than those who had attended regular classes. In the interview, the students with the intensive instruction experience produced more extended turns (defined as longer uninterrupted speech segments), and they were more accurate in the production of certain verb forms, such as past, present, or third person singular. Finally, a similar superiority was found in the Picture Card Game (one of the two communicative tasks), in which the postintensive students were more fluent (produced more words) and more accurate. In the second communicative task, in which the students asked the interviewer questions about herself, the postintensive students were usually more accurate in the use of yes/no questions. Lightbown & Spada (1991) present a detailed account on this experiment and the positive results with respect to long-term effects of the intensive English course.

In their investigation of the acquisition of oral skills in English by students in intensive courses, White & Turner (2005) used three different tasks to elicit oral production: first, the Audio-Pal (in which the students were supposed to introduce themselves to someone of their own age on an audiocassette, during approximately one minute), then, a story retell (the student retold a story

presented before in a video clip to the interviewer), and finally an Info-Gap task, for which two students interacted with each other in order to describe items which were missing in their partners' pictures. The results of these tasks demonstrated that the grade 6 students in intensive ESL made significantly more gains in all the tasks than their peers in regular English classes. This outcome could be the result of both, time concentration or the increase in instruction hours (although the latter may seem more apparent). The authors speculate that both have an effect in intensive ESL programs, and present as evidence the fact that, at the time of the pretest, the students in intensive classes were significantly superior to those in regular classes, both groups having received the same amount of hours in grade 6 prior to testing, yet such hours being concentrated in the case of the intensive group, and distributed for the regular ESL class.

An additional piece of research regarding time distribution is provided by Collins et al. (1999). As was mentioned before, there have been different models of implementation of intensive English, depending on the distribution of instruction time across the school year. Collins et al. (1999) analyzed the performance of students in *distributed*, *massed* and *massed plus* programs. In the three models, a total of 300-350 hours of instruction in English are offered, which are spread across the school year (10 months, 8 hours a week) in the *distributed* program, and five months in the two *massed* programs (18-20 hours a week). The *massed plus* model (explicitly described in Lightbown & Spada (1997) and briefly reviewed above) differs from the *massed* in the fact that in the former the students are in addition encouraged to use English outside the class (in the hallways, cafeteria, etc.). The

results from the tests that all the students took (vocabulary recognition test, listening comprehension test, and oral narrative) at the end of their course favored the students in the massed program, especially those in the massed plus (confirming results by Lightbown & Spada (1997) about the benefits of 'ambient English'). This finding seems to suggest that superior language gains are made under more concentrated second language exposure. Nonetheless, Collins et al. (1999) caption that the findings are inconclusive because the massed groups ended up having more hours than the distributed group (the former approximately 400 hours in five months, the latter about 300 hours in 10 months); moreover, the differences between massed and distributed are relative since the three groups made considerable progress from the pretest to the posttest.

In summary, the experiences in Canada with intensive language instruction (both French and English as a second language) have promoted higher proficiency levels in the case of the students who have followed them (as well as more positive attitudes and motivation towards language learning and contact with the L2 outside the class), as opposed to those students following the traditional ESL or FSL classes. Intensity in these models does not only reflect time concentration, but also an increase in instructional hours (as well as a reorientation of the curriculum, in the case of intensive French).

Whereas research being done in the Canadian context clearly demonstrates the positive effects of intensive language instruction, it does not provide much evidence for the superior effects of time concentration only, with the exception of a

few studies which suggest that concentrating instruction hours in a short period is more beneficial for the students than spreading them across time, as reflected in the studies by Collins et al. (1999), Lapkin et al. (1998), or Lightbown & Spada (1991) and Spada & Lightbown (1989), who suggest that grade 6 students following intensive English in five months acquire a proficiency level in this language at times superior to those students who had received the same amount of hours of instruction distributed over many years, who are in grade 9 or 10 (yet age or methodology used in the different types of programs could be intervening variables).

In the next section a different type of intensive program will be reviewed, which combines intensive classroom instruction with massive practice using the L2 outside of class. Such programs, called *study abroad* or *stay abroad*, generally comprise university students staying in a foreign country for a period of time in which they not only receive language lessons but also the possibility to be exposed to the L2 and practice the language in real-life situations. Comparing this experience with the intensive programs reviewed until now, it could be said that *study abroad* programs resemble the *massed plus* English intensive program in Canada (Collins et al., 1999; Lightbown & Spada, 1997), or experiences with intensive instruction in summer programs before the 1990s in which the students remained in residencies for some weeks during the summer where the use of the L2 was promoted if not required (Currall & Kirk, 1986; Keilstrup, 1981; McKee, 1983; Schneider, 1977).

3.3.4. Intensive language learning in Study Abroad programs

The experience of learning a foreign language in the Study Abroad (SA) context differs from the traditional, or At Home (AH) programs and the domestic intensive programs in a variety of ways. The most obvious difference is the access to native speakers as interlocutors, which is fairly easy and practically unavoidable in the SA context, whereas it is quite hard and virtually non-existent in the other two program types (with the exception of the instructor). Another difference, related to the topic of this research study is the concentrated exposure to the L2 that the learners experience. In a period of typically one (sometimes two) semesters, the students receive a massive amount of hours of contact with the foreign language both through classroom instruction and out of class.

However, the fact that contact with the L2 input is easily accessible does not mean that the students will take advantage of such availability. Segalowitz & Freed (2004: 196) present the claim clearly: "Contexts differ in terms of what learning opportunities they present. Learners differ in terms of how ready they are linguistically and cognitively to seize the opportunities provided and to benefit from them once they do." Individual variation is therefore an issue which must be considered in SA, since there are many intervening factors in the acquisition of the L2 in such a context; not surprisingly such variability has been claimed to be higher for foreign language learning in SA than in AH programs (Freed, 1995; Regan, 1995). One individual characteristic that may affect the acquisition of an L2 in a SA context may be the student's personality, which will determine how much

contact with the natives is sought (DeKeyser, 1991). Gender may also be a variable which affects language acquisition in the SA context. Some studies have reported that females make less progress than their male counterparts in countries where the women's role is highly different from the L1 country, as is the case of American women in Russia (Brecht, Davidson, & Ginsberg, 1995; Polanyi, 1995). These studies report less language gains for women since they could not interact as freely as the males in their program, due to the sexual harassment experienced by some of them in the foreign country. Other individual variables which may affect acquisition in a SA context include previous foreign language learning experience, which is supposed to be a facilitating factor in L2 learning in SA (Brecht et al., 1995). Additionally, the students' age can also be an influential factor. Brecht et al. (1995), in their study which analyzes the performance of Americans learning Russian in Russia over many years, report that younger learners tend to make more progress than older learners; furthermore, the authors include age as one of the predictors of foreign language gains in SA.

Another predictor of L2 gains in a SA context is initial L2 proficiency level. Several studies have found that it is harder for advanced learners to make progress after the SA experience than it is for intermediate learners[28]. Brecht et al. (1995) in their study of the acquisition of Russian by American students found that those with the higher initial level of Russian were less likely to make gains after staying abroad. Similarly, Lapkin, Hart, & Swain (1995) observed that, in an

[28] These findings are not conclusive, since opposite results have also been obtained (see DeKeyser, 2007b).

'interprovincial' exchange in which English-speaking Canadian adolescents (grades 10-12) spent a period of time in Quebec in order to boost their knowledge of French, the students who made the most gains after the stay (especially in terms of listening and oral skills) were those whose scores in the pretest were lower. Additionally, in her analysis of fluency in French in the SA context, Freed (1995) reported that those students that were rated as less fluent before the SA experience were perceived as having made the most progress in this area in the posttest. In an earlier study, Freed (1990) also reported that the students who studied French abroad during 6 weeks who started with an intermediate proficiency level made more progress in their French skills than those who were more advanced. Moreover, she claimed that interactive out-of-class contact helped the students improve their French skills more than non-interactive contact (exposure to French through the media, or other passive means). Lower proficiency students have also been claimed to demonstrate a higher improvement in vocabulary acquisition than advanced learners in the SA context (Milton & Meara, 1995). Similar findings were obtained for sociolinguistic competence, in which advanced learners were not reported to make many gains after some time in the L2 country (Regan, 1995; 1998). Therefore, although some studies have suggested that there is no difference in foreign language learning in a SA context in terms of the students' initial L2 proficiency level (Ife, Vives-Boix, & Meara, 2000 for vocabulary), there seems to be a general agreement that the students who benefit the most from a SA experience are those who, already having a certain command of the L2, are not advanced learners. Some authors have pointed out that the reason for this apparent lack of

progress at the advanced level lies on the type of tests that has been used. As suggested in Freed (1998) the OPI (which has been frequently employed to analyze oral data) cannot discriminate well between the students with advanced proficiency because of its non-linear construction. Ife et al. (1998) argue that measuring can be problematic at the upper levels if improvement in the L2 is analyzed as percentage of items acquired (for instance in the case of vocabulary): the more knowledge a student has, the more difficult it will be for him or her to make proportionate gains.

With respect to the overall results of students in language tests after the SA experience (which is usually one semester, although some studies present data from learners who were abroad for two semesters), research in the area suggests that greater language gains are made in this context as compared to the AH programs (which typically offer from 2 to 4 hours of foreign language instruction a week). Such gains have most often been claimed in the domain of oral fluency, defined as speech rate and length of speech between pauses (Möhle, 1984), or in terms of speech rate, mean length of speech run not containing filled pauses, and longest fluent run not containing silent hesitations or filled pauses (Segalowitz & Freed, 2004). Other studies which have reported greater improvements in the SA group with respect to oral fluency include DeKeyser (1991), and Freed (1995). Lafford (2004) also suggests that students are more fluent after staying abroad due to the fact that they use fewer disruptions through communication strategies. In terms of written fluency, not much improvement has been claimed for the students in the SA context (Freed, 1998).

The progress SA students make in other language areas, apart from fluency, has not been usually reported to be significantly superior to AH students. Díaz-Campos (2004) did not observe much difference between SA and AH students in the acquisition of certain phonological features in Spanish. Similarly, Dewey (2004) did not claim superior reading comprehension skills in Japanese by the SA students as opposed to those staying at home. In terms of grammar, there is not much evidence supporting the advantage of students in SA programs as opposed to those in the AH context. Whereas there are some studies that report on higher morphosyntactic gains by students abroad (Lennon, 1990), most research studies have failed to find such advantage (Collentine, 2004; DeKeyser, 1991; Regan, 1995). On the other hand, students in the SA context have been often reported to significantly increase their vocabulary after their experience in the foreign country (DeKeyser, 1991; Ife et al., 1998; Lennon, 1990; Milton & Meara, 1995).

What many studies analyzing the effects of the SA experience on L2 learners have claimed is that most educators and researchers perceive that the majority of the students after staying abroad demonstrate a qualitative change in their L2 skills; nevertheless, the measures which have traditionally been used to analyze their progress (and compare it with learners in the AH context) tend to focus on features which are highly related to formal instruction; that is why many studies have found advantages for the AH context (Collentine, 2004). It is important that measures which examine other types of language gains are developed in order to quantify the impression that "the SA learner can 'tell a story' a little better and can 'get their point across' more effectively" (Collentine, 2004:

245).

It is true that there are some gains in the SA context that are harder to quantify, yet most (if not all) the students staying abroad also received formal instruction even in higher amounts than the students in the AH context. It is thus surprising that the SA students' results in areas related to grammar complexity or vocabulary are not superior (or are in fact lower in many cases) with respect to their peers at home (Collentine, 2004), since they had the advantage of both classroom instruction and out-of-class contact with the language. One explanation can be that the gains in fluency which are unarguably attributed to students in SA contexts are made at the expense of growth in other areas, such as grammar complexity or accuracy.

DeKeyser (2007b) suggests that the main reason why the students in the SA context do not make as much progress as expected in their language skills is the fact that learners do not get appropriate L2 practice. According to this author, when students go abroad they should have a procedural knowledge of the foreign language, and the time abroad should coincide with the stage of automatization. The problem that many students face in the SA context is that they have not had a chance to proceduralize their declarative knowledge of the L2 in their classes before being exposed to the foreign language in real-life situations. Often, L2 classroom practice does not resemble every day communication and, as a consequence, it does not provide opportunities for the students to proceduralize their declarative knowledge of the L2. When these learners are in a SA context, they are so overwhelmed by the communicative demands which they encounter

that instead of proceduralizing their knowledge of the L2, they simply acquire new declarative knowledge through formulas, which they automatize through constant practice. Therefore, the learners are not automatizing rules, but formulas, which is why their overall performance has been found to be less than optimal.

Although most comparison studies with learners in the SA context have taken place with regular AH programs, some research has been done comparing the students' gains in SA, AH, and domestic intensive courses. Surprisingly, when investigating fluency, Freed, Segalowitz, & Dewey (2004) discovered that the students in the intensive program (receiving 7 weeks of French instruction during the summer, with approximately 17.5 hours a week) made significant gains in the total number of words, length of the longest run, rate of speech and speech fluidity, whereas the SA students made gains only with respect to speech fluidity and less significant than those experienced by the students in the intensive program. The learners in the AH program did not make any significant gain according to the fluency measures used in this study. When examining the data obtained from the out-of-class contact questionnaire, it was evident that, thanks to the large number of extracurricular activities organized for the students in the intensive domestic program, those learners reported to have used more the L2 than their peers in the other two program types.

Another study comparing the development of reading comprehension in Japanese by students in SA and domestic intensive (Dewey, 2004) found no significant differences between the two contexts, except for self-assessment: the students in the SA context felt more confident of their reading abilities than those

in the intensive 'at home' program. From these results Dewey (2004) concludes that a 9-week intensive summer course can produce gains in reading abilities as determined by objective reading measures comparable to an 11-12 week stay in Japan.

In summary, it can be said that the progress made by the students in the SA context tends to be lower than what is expected, considering the amount of input received and the ample opportunities for output practice that the students have at their disposal. Nevertheless, individual variables, as well as initial proficiency level and stage of language acquisition are determining factors which can account for L2 development in the SA context. Most studies analyzing gains experienced by students abroad tend to focus on oral proficiency skills, and oral fluency is the area in which more advantages have been reported for SA students as opposed to those who stay at home. Another important finding which must be considered from research in the SA context is that, in some cases, intensive at home programs have been shown to be more effective in promoting students' L2 skills than a period abroad.

3.3.5. Time distribution and language learning: Some evidence from the Spanish context

There has not been much research on the issue of how distributing the hours of foreign language instruction affects foreign language development in the Spanish context. Nevertheless, there is some indirect evidence in favor of

concentrating the hours of L2 instruction from studies whose main objective was to provide an answer to other research questions unrelated to intensity of L2 exposure.

When investigating the age factor in foreign language acquisition, several authors have demonstrated that those learners who started learning English when they were older (11 years old) had superior writing skills when compared to those who started learning this language when they were 8 years old (Celaya, Torras, & Pérez-Vidal, 2001; Torras et al., 2006). Apart from the age difference between the two groups of learners, there was a difference in the concentration of instruction hours, with those learners who began learning English at a later age receiving more concentrated language exposure. Torras et al. (2006) suggest that such concentration may have had a positive effect on the acquisition of L2 writing skills by the late starters.

Additionally, several reports published by the Spanish Education Ministry (Ministerio de Educación y Ciencia, MEC) claim that the students who spend more hours a week practicing English show a more advanced performance in the language, as compared to those students whose contact with the language is more limited. The report on the teaching and learning of the English language in the year 2000 referring to the evaluation performed in 1999 in primary education states that those students who received more than three hours of instruction a week were significantly superior to those who received fewer in all the skills evaluated (listening and reading comprehension, and written production). Furthermore, advantages are reported for those students who practiced their English skills

outside the class in all the language areas.

Similarly, the MEC 2004 report referring to secondary education in 2001, confirms significant advantages in English skills for those students who practiced the language outside the class (which obviously means more practice). Improved performance can be observed for those students who stayed abroad for a period of time, and for those who practiced the language with friends or relatives. More importantly, and as was reported for the students in primary school, those learners in secondary school who received more hours of English instruction a week (four or more) show significantly more skillful performance in the language than those receiving two or three[29].

From the two MEC reports we could interpret that when exposure to the L2 is limited and distributed in a 'drip-feed' fashion, the effect of such exposure or hours of instruction is quite limited in terms of the students' improvement of English language skills. On the other hand, those students who establish more intensive contact with the language (through more hours of practice or concentration of instruction hours) demonstrate a superior performance in most language skills.

Research on Content and Language Integrated Learning (CLIL) in the Spanish context has presented some advantages in terms of English language proficiency for the students who have received this type of instruction as opposed

[29] There is no specific reference on this report regarding the total amount of hours of English instruction the students had by the end of secondary school, depending on whether they had received two, three or four or more hours of class every week. It is only said that the classes varied in duration from 45 to 60 minutes, yet no specific information is given regarding whether more classes a week meant an increased number of hours of instruction.

to students in regular classes (Muñoz, 2004). The reason for such advantage can be related to the optimal acquisition of language through implicit learning when the students are focusing on content, but also to the fact that there is increased and more intensive contact with English in CLIL than in regular foreign language programs.

In general, it can be said that the evidence from studies in the Spanish context investigating the acquisition of English in schools tends to suggest that, under different circumstances, L2 programs which have provided the students with a higher amount of instruction hours or have concentrated the hours of EFL instruction tend to show certain advantages with respect to other programs in which the exposure to the foreign language is either reduced or more spread out in time.

3.4. Summary of Chapter 3

This chapter has analyzed the role of time distribution of practice in learning in cognitive psychology as well as time distribution of instruction hours in SLA. Section 3.1 has presented some evidence from the cognitive psychology literature regarding the superior effects of distributed practice for long-term retention of to-be-learned material. First of all, it was argued that contextual variability between P1 and P2, when such presentations are spaced, facilitates the learning of the repeated items. Another explanation for the spacing effect which was reviewed assumes that massed presentations are not as thoroughly processed

as spaced presentations, due to the fact that the recency of P1 at the time of P2 makes less processing necessary for the retrieval of P1. Additionally, it was argued that even if accessibility of P1 at the time of P2 should be made difficult in order for deeper processing to occur, P1 needs to be retrievable for P2 to strengthen such presentation. Distributions which are too spaced hinder retrieval and thus have negative effects for learning and retention. Another important finding from the experiments in cognitive psychology is that massed presentations, due to recency effects, are better remembered at immediate tests, but their retention is worse at delayed tests than the material presented in spaced sequences.

Sections 3.2 and 3.3 illustrate how concentrating instruction hours can have positive effects for learning in school contexts in high school or college in the case of subjects other than foreign languages, but especially for L2 acquisition. Studies analyzing block scheduling in high school and accelerated courses at university have reported advantages for such educational models, since they encourage deeper discussion and more time-on-task. Similarly, research on intensive language instruction has demonstrated that concentrating (and also increasing) the hours of L2 instruction is highly positive, especially for improving students' speaking skills. Such positive findings have been reported for adult learners as well as for school children, the latter mainly in the Canadian context.

The results from the experiments reported in the cognitive psychology and in the SLA literature may lead one to believe that the evidence they provide is contradictory. Nevertheless, the different nature of the experiments, as well as the different type of knowledge that is supposed to be acquired in each context can

explain the dissimilarity in results. Language teaching as is implemented in most classrooms today hardly ever involves constant and continuous repetitions of the same patterns or words (which would correspond to massed practice in the cognitive psychology experiments). There are repetitions of grammar structures, vocabulary items or collocations in the input that foreign language learners receive, yet such repetitions always appear in distributed sequences, with other material intervening in the presentations of the target items; moreover, repetitions appear in different contexts. As a consequence, both in intensive as well as in traditional programs language learning is distributed. In intensive programs, though, the repetitions of the declarative knowledge that the students are supposed to acquire occur in less spaced sequences, which makes such declarative knowledge more available for retrieval and thus for proceduralization. When the duration of foreign language lessons is limited, and long periods of time elapse between sessions (as in most L2 classes), the students may have problems retrieving the declarative knowledge previously acquired, which would make proceduralization harder. Therefore, taking into account the predictions made from accessibility theories as well as study phase retrieval theories of the spacing effect, intensive instruction should have more positive effects on L2 skills than *drip-feed* classes.

CHAPTER 4: RESEARCH QUESTIONS

The aim of this dissertation is to shed some light on the effect of time distribution on the acquisition of English as a foreign language for students with different initial levels of proficiency. The issue of how the hours of instruction should be distributed in foreign language learning has not received much attention in the last few years, except for the large-scale studies which have been performed in Canada since the 1990s for intensive French (Germain, Netten, & Movassant, 2004; Germain, Netten, & Séguin, 2004), but mainly for intensive English (Collins et al., 1999; Lightbown & Spada, 1991; 1994; Spada & Lightbown, 1989; White & Turner, 2005). For adults, there is a dearth of studies on time distribution which present comparison data with students in intensive and regular foreign language classes (Serrano & Muñoz, 2007), and even more so comparing different proficiency levels. This particular study strives to provide empirical evidence for the optimal time distribution of instruction hours according to the students' initial level of proficiency in English. More specifically, the research questions which guide this study are the following:

Research Question 1: When the total amount of hours of instruction is held constant, does the distribution of such hours (manifested in extensive, semi-intensive, and intensive English courses) have any effect on the acquisition of English as a foreign language, as reflected in students' gains in listening, grammar, vocabulary, reading, writing, and speaking skills?

It was examined in Chapter 3 that some empirical studies have demonstrated the positive outcomes of students in intensive language programs as compared to their peers registered in traditional foreign language classes. Most of those studies referred to young learners (primary/secondary school) who were enrolled in an intensive language program that not only concentrated the hours of instruction but also increased them, and the positive outcome of such programs was observed mainly in the students' listening, reading and speaking skills (Collins et al., 1999; Lightbown & Spada, 1994; Spada & Lightbown, 1989), but also in their vocabulary (Collins et al., 1999). Fewer studies have analyzed gains in writing skills; nevertheless, Lapkin et al. (1998) reported higher gains in writing and also reading skills for students receiving more concentrated hours of instruction in primary school. Similarly, McKee (1983) found that college students in an intensive French program showed more advanced writing skills than their peers in semester-length courses. Even if no empirical data has been offered in terms of language improvement for students in semi-intensive programs, some authors have already suggested that such programs have a positive effect on students' language learning (Cipolla, 1982; Oswald, 1950).

Taking into account the findings from the cognitive psychology literature with respect to the fact that previous presentations of repeated material have to be accessible when such material reappears, together with Anderson's theory of cognitive skill acquisition (Anderson, 1993) and DeKeyser's (2007a; 2007b) adaptation ACT-R for language acquisition, it could be postulated that the reason why intensive programs promote higher levels of language acquisition is due to

the more accessible declarative knowledge that the students in such programs have available at the time proceduralization takes place. Therefore, considering the evidence from both the cognitive psychology and the language program evaluation literature, it could be hypothesized that the learners included in this research who followed intensive language instruction (both intensive and semi-intensive) should have an advantage in the acquisition of listening, grammar, vocabulary, reading, writing and speaking skills in English with respect to their peers in non-intensive classes.

Research Question 2: Does time distribution have a different effect on students at the intermediate and advanced proficiency level in terms of the language gains experienced by those students at the end of their respective course?

To this researcher's knowledge, no study has analyzed the effect of time distribution on L2 learning that considers the students' initial proficiency level by providing comparison data from learners with a different command of the L2 in intensive and non-intensive language courses. Most studies which have investigated the effect of time distribution on language learning have included students who were at beginning or intermediate stages in their acquisition of the foreign language. When the performance of such students was compared to other learners who had been exposed to less intensive instruction, more gains became evident in the case of those students attending concentrated L2 courses (Collins et al., 1999; Lapkin et al., 1998; Lightbown & Spada, 1994; McKee, 1983; Spada &

Lightbown, 1989, White & Turner, 2005). Consequently, similar gains should be expected for the intermediate (level 3) students enrolled in the intensive program included in this research study.

Studies investigating the effect of time distribution on advanced language learners, comparing the performance of students in intensive and non-intensive language courses at this level, are practically absent. There are some researchers, however, who have analyzed learners' gains in intensive programs depending on the students' initial proficiency level (Gardner et al., 1977; Lapkin et al., 1995). Such studies tend to report that the lower the learners' proficiency level is at the beginning of the course, the more evident the language gains are when the course is completed.

Considering the power law of practice examined in section 2.1, it can be expected that less improvement occurs at later acquisition stages. Thus, it could be hypothesized that different effects should occur for the distribution of instruction hours in the case of intermediate and advanced EFL students, bearing in mind the different stages in the language acquisition process in which the learners at the two levels under study are. Previous research on intensive language courses has shown that concentrated language exposure facilitates proceduralization of language skills (by making declarative knowledge more available for retrieval and practice), since the performance of students in intensive groups at the intermediate stage of L2 acquisition seems superior to that of students who do not receive concentrated exposure. When students are at a more advanced level of proficiency, what they can learn in foreign language classes are more specific or elaborate applications of

the rules that they already know and have proceduralized. Similarly, foreign language instruction at the advanced level might also be necessary to eradicate fossilized interlanguage forms. It seems, therefore, that, providing the learners under examination are at different acquisition stages and they are expected to acquire different aspects of the L2, time distribution may also affect these two groups of learners differently.

CHAPTER 5: METHOD

5.1. Program and participants

Three different programs were chosen for this study: extensive, semi-intensive and intensive. These programs were offered in the same institution, which is the language school of the University of Barcelona. This language school has been working for more than 50 years and it offers a wide selection of foreign languages, with English and French being the most popular ones. With respect to the English language, courses are offered from the beginner level to proficiency, distributed in six different levels. The correspondence of the schools levels with the levels established by the Common European Framework of Reference for Languages (Council of Europe, 2001) is the following:

- Level 1: Basic User (A1)
- Level 2: Basic User (A2)
- Level 3: Independent User (B1)
- Level 4: Independent User (B2.1)
- Level 5: Advanced User (B2.2)
- Level 6: Advanced User (C1)

This dissertation focuses on level 3 (intermediate) and level 5 (advanced) courses of EFL[30]. As mentioned above, the level 3 corresponds to the Common European Reference Level B1. The learners at this level, according to the Common European Framework of Reference for Languages (p. 24), can do the following in the L2:

> Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.

Level 5 corresponds to level B2.2 of the Common Reference Levels, which would be between levels B2 and C1 (p. 24). The learners at these levels can do the following in the L2:

> B2: Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
>
> C1: Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously

[30] The original purpose of the study was to include level 1 (beginners) and level 4 (upper-intermediate) students too; in fact, data was collected from some groups; nevertheless, due to the fact that some teachers in such levels were not interested in participating in this research project, those levels finally had to be excluded from the analyses, since not enough data could be gathered. However, the students' performance in levels 3 and 5 is expected to be different enough to produce interesting results regarding the effects of time distribution on EFL with respect to the students' initial proficiency level.

> without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices[31].

Three different programs were chosen for this study: extensive (for both, intermediate and advanced students), semi-intensive (only for intermediate students, since such courses are only offered for level 3 and level 4), and intensive (for advanced and intermediate students). In all the three models English is taught for 110 hours; however, those hours are distributed in a different way. Extensive programs start in October and they offer four hours a week of English instruction distributed over two days, Mondays-Wednesdays or Tuesdays-Thursdays, in which the students receive two hours of instruction a day (there are usually 10-minute breaks in each session). These classes continue until the end of the school year in May, over a total of approximately seven months.

Semi-intensive courses are offered twice over a school year: during the first semester from October until December, and during the second semester from February until May. As mentioned before, only students in levels 3 and 4 can follow such courses. Students enrolled in the first semester semi-intensive level 3 receive ten hours of instruction a week (two and a half hours a day Monday through Thursday) over a total of approximately eleven weeks. On the other hand, those students attending the semi-intensive course in the second semester receive eight hours of instruction a week (two hours a day Monday through Thursday)

[31] For more detailed information about the Common Reference Levels and their characterization see Council of Europe (2001).

during approximately fifteen weeks[32]. As happened in the case of extensive courses, there are also 10-15 minute breaks in semi-intensive lessons.

Finally, intensive courses are offered for levels 1 through 5 during the summer. These courses provide five hours a day of English instruction from Monday through Friday during a period of nearly five weeks from the end of June until the end of July. There are usually two breaks of 10-15 minutes in each session; however, the time allotted varies depending on the instructor or the day (although it is less typical, some teachers decide to have only one longer break).

The reason why these programs are offered at this institution is not related with previous research studies on the efficiency of intensive programs, but with students' usual lack of available time during the school year to take English lessons. The semi-intensive and intensive classes allow the students to get the free election credits they need to complete their degree (or to learn English, depending on the students' motivations) in a shorter time and with a lower probability of a schedule clash with the obligatory subjects within their respective degrees.

The number of participants distributed by level, and program type are presented in the following table (Table 5.1):

[32] In the first and second semesters the distribution of instruction hours in the semi-intensive group is slightly different. However, that is the design adopted by the school every year; moreover, the differences between the two semi-intensive formats are narrow and both can still be clearly differentiated from extensive and intensive.

Table 5.1 Participants

	Students	Groups	Teachers
Level 3 Extensive	49	4	4
Level 3 Semi-Int.	44	3	3
Level 3 Intensive	38	3	3
Level 5 Extensive	34	4	3
Level 5 Intensive	31	3	3
TOTAL	**196**	**17**	**13**[33]

Approximately half of these groups were included in the academic year 2004-2005 (two level 3 extensive, two level 3 semi-intensive, two level 3 intensive, two level 5 extensive and two level 5 intensive), and the remaining groups (eight out of the seventeen included in this study) attended classes during the academic year 2005-2006. Regarding the number of students, there are many more included in the academic year 2004-2005 than in the following academic year (112 vs. 85). The reason for this difference is that the objective of second year of data collection was to have a large enough group for each level in each type of program; consequently, the data collection did not aim at having a sample as large as the previous year. Additionally, the data collection process became more and more difficult, as there were more and more teachers unwilling to participate.

A point that should be made is that the number of students registered in each group was superior to the number of participants included in the study, since not all students who were registered in a specific group took the tests. The following table provides more details with respect to groups and students who participated in this study.

[33] The total number of teachers is not the addition of the teachers in each group, since there are four teachers who taught more than one group.

Table 5.2: Groups

	Group	Academic year	Students enrolled	Students in study	Teacher
Extensive	**3J**	2004-2004	20	9	A (NSS)
	3V	2004-2004	21	13	B (NES)
	3L	2005-2006	24	11	C (NES)
	3X	2005-2006	24	16	D (NSS)
Semi-intensive	**3Semi1**	2004-2004	19	16	A (NSS)
	3Semi2	2004-2004	22	17	E (NCS)
	3Semi3	2005-2006	19	11	F (NES)
Intensive	**3IntA**	2004-2004	18	14	G (NES)
	3IntB	2004-2004	9	7	H (NES)
	3IntC	2005-2006	23	17	C (NES)
Extensive	**5G**	2004-2004	11	9	I (NES)
	5H	2004-2004	15	12	I (NES)
	5I	2005-2006	13	6	J (NES)
	5N	2005-2006	13	8	K (NES)
Intensive	**5IntA**	2004-2004	11	9	F (NES)
	5IntB	2004-2004	9	7	L (NES)
	5IntC	2005-2006	18	15	M (NES)

NSS: Native Spanish Speaker, NES: Native English Speaker, NCS: Native Catalan Speaker

Many students in each of the groups under research did not participate in this project. The reasons for this can be found in the data collection procedure, which included pre and posttest (the procedure will be explained in more detail in section 5.2), but also in the inconsistency of class attendance, which is common for a lot of students especially in extensive groups.

One point that should be mentioned is that the numbers of students enrolled in extensive groups was generally higher than in semi-intensive or

intensive groups. At the same time, attendance is less regular in whole-year courses (see Figure 5.15 and Figure 5.33 for more details); therefore, the number of students who were actually in class was not extremely different in the types of programs included in this dissertation. Another issue to consider is that the number of students in level 5 classes is almost always lower than in level 3 classes, especially in extensive groups (the average number of students in the level 5 classes under study who actually came to class was 8-10, whereas for the extensive level 3 groups was 15-18).

The students who attended the classes in which the data were collected were similar across program types. All of them were adults (18 years or older), with a high school degree or a more advanced diploma, native Spanish/Catalan speakers (almost all of them bilingual) and residents of Catalonia, Spain. Figure 5.1 through Figure 5.6 present some graphs with background information about the students in **level 3**, which was obtained by means of a questionnaire, distributed at the end of the course. It must be mentioned that not all the students had time to complete this questionnaire.

It can be observed that the three programs under examination are quite similar in terms of students' age. Most of the students in the three programs fell within the 18-23 year-old range, which are typically university years. The only difference that can be mentioned is that, in general, the population of the intensive groups is slightly younger, since there are no participants who were older than 35. In terms of gender (Figure 5.2), there are more females than males in this program type (as well as in the extensive course), yet in the semi-intensive program there

are slightly more males. Figure 5.3 confirms that most of the participants in all the groups are university students who either just study (the majority in the three program types), or work and study (W&S). The distribution in terms of students' occupation is highly similar across groups, as is their education level (Figure 5.4), with most students having a high school degree (HS), few of them having obtained their bachelor's degree (BA/BS), and even fewer pursuing graduate studies (Grad). When analyzing the participants' study field, it was obvious that for the three program types most students were doing scientific degrees (Sci) (see Figure 5.5). Figure 5.6 represents the students' reported knowledge of languages other than Spanish, Catalan and English. Most students' language knowledge is limited to these three, yet there are some students who claimed to have some knowledge of one more language (1) (usually French), and very few two (2).

Figure 5.1: Participants Age Level 3

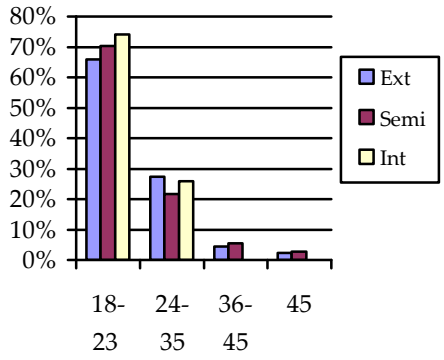


Figure 5.2: Participants Gender Level 3

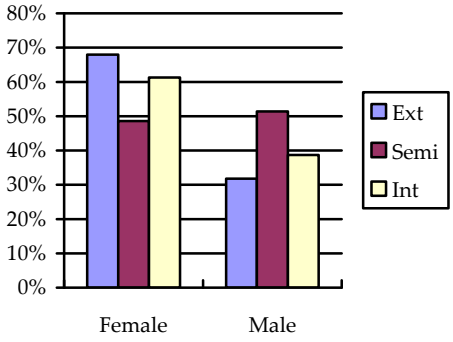


Figure 5.3: Particip. Occupation Level 3

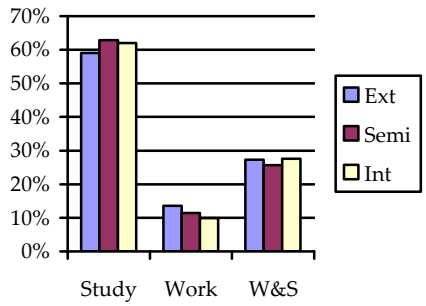


Figure 5.4: Particip. Education Level 3

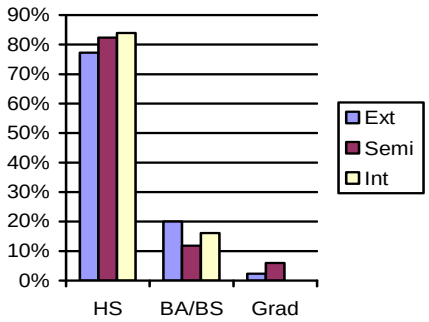


Figure 5.5: Particip. Study Field Level 3

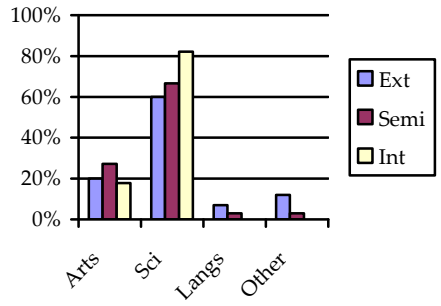
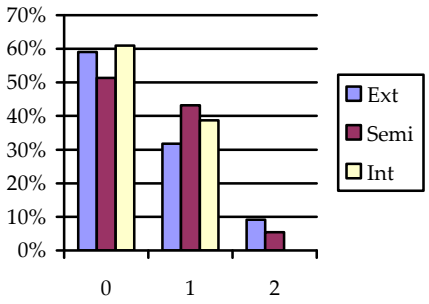


Figure 5.6: Particip. Other Lang Level 3



Comparing the three programs, again, they are quite similar; the only difference being that in the intensive program there were no students who knew two

additional languages apart from Spanish, Catalan and English.

It was also interesting to find out about the students' experiences with the English language, especially about how many years they had been studying it, when they had started and how continuous the exposure had been. Moreover, the students' motivation was examined with questions like "Do you like English?" or "Why are you studying English?" The answers to these questions are represented in Figure 5.7 - Figure 5.12. In terms of the amount of years studying English, it can be observed in Figure 5.7 that whereas the majority of the students in extensive and intensive groups fell in the >10 (more than 10 years) range, in the semi-intensive most students had studied the language between 5 and 10 years (which probably indicates that they did not take any English class after high school). Very few students have been learning English for 1-5 years. Those participants claiming to have studied the language for less than 2 years could have only done so if they did not study the language in high school and have taken intensive or semi-intensive courses (since it is difficult to attain an intermediate level in less than 3 years in regular courses). Most students in the three program types started learning English when they were 6-10 years old (see Figure 5.8), with a high number also falling in the 11-15 range (these two ranges would reflect the educational change in the 1990s which moved the starting age for foreign languages at school from 11 to 6 years old (Muñoz, 2006). In terms of continuity studying English (Figure 5.9 and Figure 5.10), students in the semi and intensive programs have had a slightly more discontinuous exposure to the language, as reflected by the number of participants in the "Discont.", and also taking into

account that there is a higher number of students in these two program types whose previous English class had taken place four or more years before (Figure 5.10). When examining the students' motivation, it can be observed that there are no differences in the three programs under research, since most students in the three of them picked ´4´ in a five point scale (from ´little´ to ´a lot´) when asked about whether they liked English (see Figure 5.11). Figure 5.12 shows that, despite the fact that most students seem to like English, the reason why they were doing an English course was because they needed such class, the majority of them for completing their degree, since they were taking the subject to obtain 'free election' credits. Others reported that they needed it for their job (present or future), while the 'Oth' or 'Mix' columns represent those students that took the class because they needed the language, but also because they liked it, or more than liking the language itself, they enjoyed being able to communicate with many people using it, and traveling all over the world. Summarizing, it can be said that, regarding the subjects' experience with English and their motivation to learn it, the three groups are quite similar. The only slight differences that can be mentioned are that the students in the semi-intensive program had, in general, received fewer years of instruction, and had been more discontinuous in their practice (this second issue also being true for the students in intensive groups).

Figure 5.7: Years Learning Eng Level 3

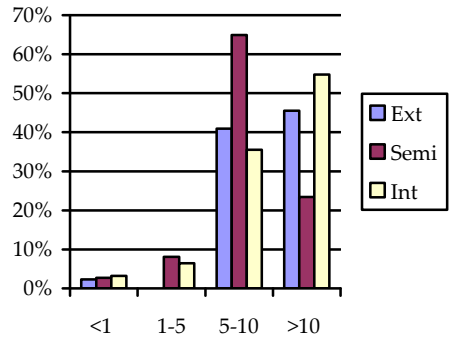


Figure 5.8: Age Started Eng Level 3

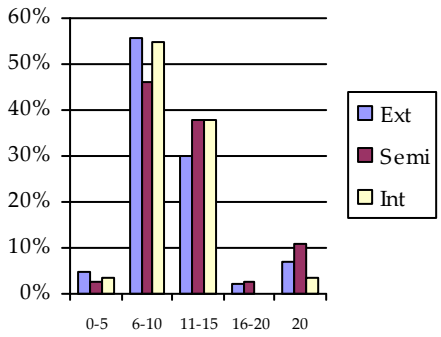


Figure 5.9: Continuity Eng Level 3

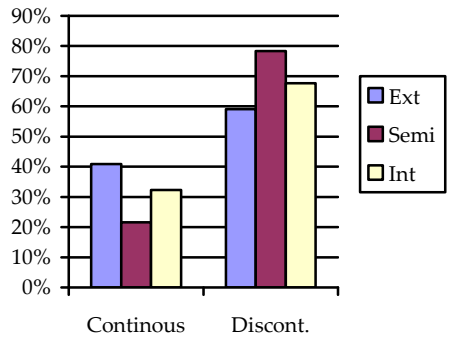


Figure 5.10: Previous Eng Course Lev. 3

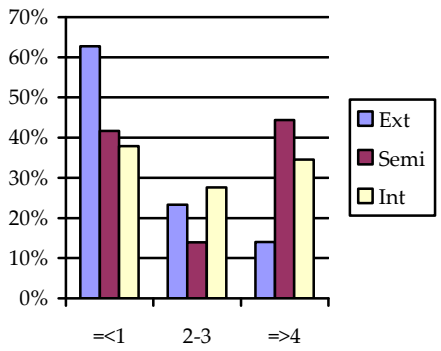


Figure 5.11: Like English Level 3

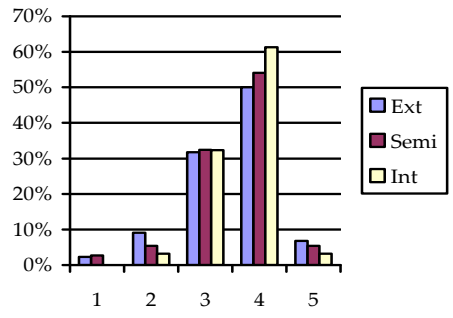
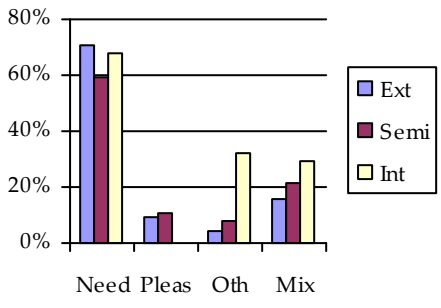


Figure 5.12: Reasons to Learn Eng Lev 3



Also related to the subjects' experience with English is whether they had been exposed to the language in a natural context. In order to examine this issue, the questionnaire included questions regarding whether the students had been in a country where they spoke English (Figure 5.13) and how long they had stayed there (Figure 5.14). In general, it can be observed that the students in the intensive program had had less natural exposure to the language than the students in the other two program types. Then, considering the extensive and the semi-intensive, the latter seems to have practiced more English abroad than the former.

Figure 5.13: Eng-Sp Country Level 3

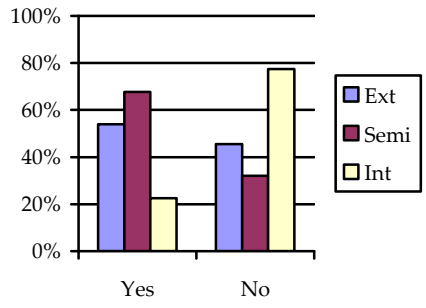


Figure 5.14: Time Eng-Sp Count Lev 3

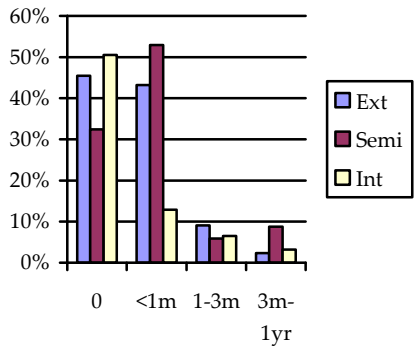


Figure 5.15 to Figure 5.18 illustrate some aspects related to the course where the students were registered at the moment. It can be seen in Figure 5.15 that attendance was much higher for students in intensive groups than in extensive or semi-intensive, since practically all the students in the former attended 80-100% of the classes, which can be expected, since missing one class in the extensive program means missing two hours, in the semi-intensive two and a half or two

(depending on the semester), and in the intensive five. Figure 5.16 shows that whereas most students in the three program types thought that they had learned 'Ok' or 'a lot', most students in the intensive groups thought that they had learned 'a lot', some even 'really a lot', while for the other two program types 'Ok' was the most common response. The students were also asked which language area/s was/were devoted more time in class: a) grammar/vocabulary, b) speaking, c) reading, d) listening, e) writing, and f) other. Most students in the three programs picked several options or the 'other' option, which usually referred to 'quite balanced' (all the different skills being given a similar weight), or 'mixture of several' (see 'Mix' column, Figure 5.17). A few students picked grammar/vocabulary (G/V) or 'Sp' (speaking) as the skills that were more practiced in class, again with similar results across groups. Therefore, it can be said that according to the students' perceptions, the classes in the three program types were quite balanced and the emphasis given to the different language skills was similar across groups. Figure 5.18 illustrates what can be expected: that the more hours the course has a week (or the more intensive the course is), the more hours doing homework the students spend.

Figure 5.15: Attendance Level 3

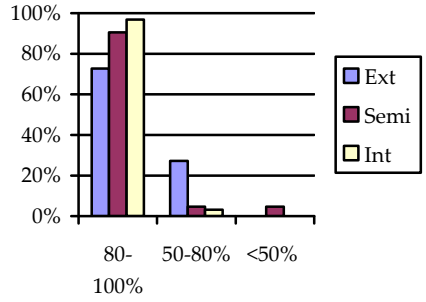


Figure 5.16: Learned this Year Level 3

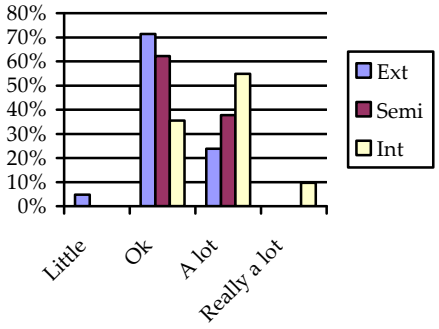


Figure 5.17: Classtime Lang Areas Lev 3

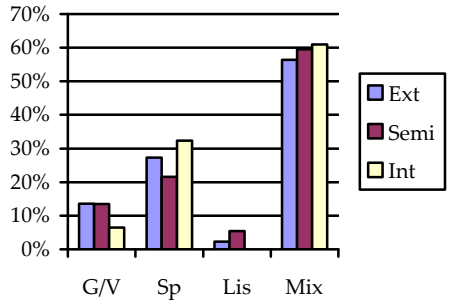
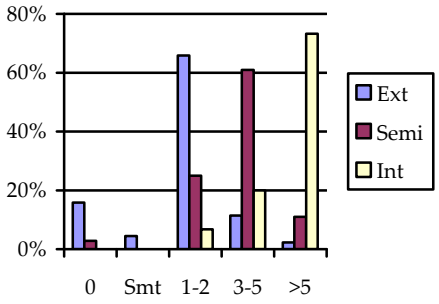


Figure 5.18: Hours/Wk Homewk Lev 3



With respect to the **level 5** students, there was, in general, a younger population than for level 3, since there were no students older than 35 years old (see Figure 5.19). Most of the students in the extensive program were between 18 and 23 years old (typical age for university students), and a few of them between 24 and 35. In the case of the intensive groups, the distribution was more balanced with respect to these two age groups, with slightly more students who were 24-35 years old. In terms of gender, there were more females than males, yet, the distribution was more balanced in the intensive program (see Figure 5.20). As was

the case for the level 3 students, in level 5 most participants studied, or worked and studied (W&S) at the same time (Figure 5.21). In level 5 extensive the majority of the students only studied. However, in the intensive program most students worked and studied at the same time, which probably explains why they chose the intensive course: most of them did not have time to take English classes during the academic year. In terms of education, most participants in level 5 had a high school diploma and were completing their university degree, which most commonly was a scientific degree (see Figure 5.23, column 'Sci'). Figure 5.24 shows that approximately half of the students in the intensive and extensive program knew one more language apart from Spanish, Catalan and English. There were few students who knew two more languages, and those who did were in the extensive program.

Figure 5.19: Participants Age Level 5

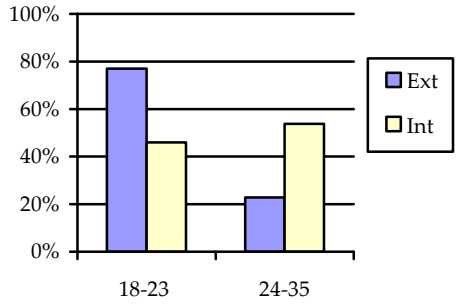


Figure 5.20: Participants Gender Level 5

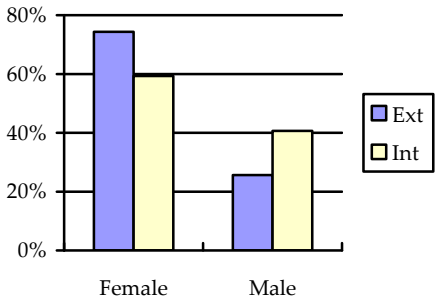


Figure 5.21: Particip. Occupation Lev 5

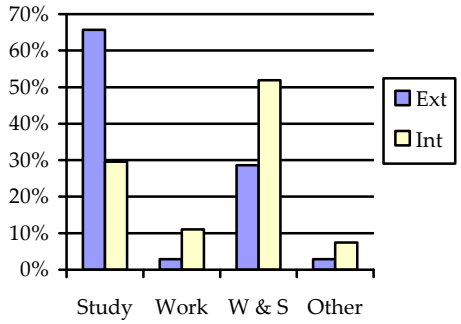


Figure 5.22: Particip. Education Level 5

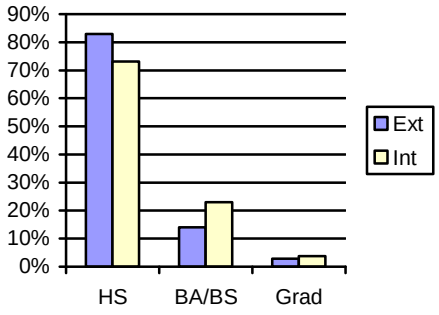


Figure 5.23: Particip. Study Field Lev 5

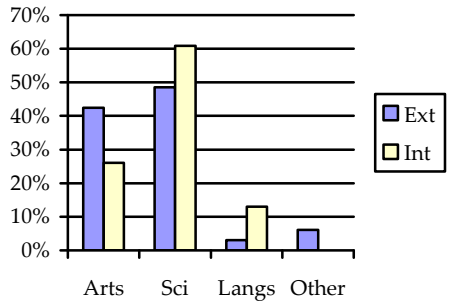
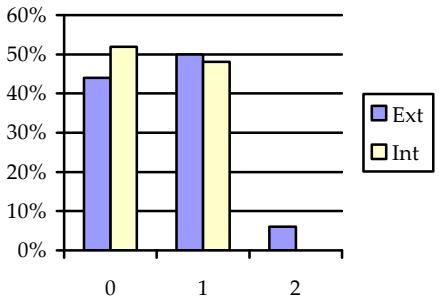


Figure 5.24: Particip. Other Lang Lev 5



Regarding the level 5 students' experience with the English language, most of them had studied it for longer than 10 years. There were a few students who had studied it for less than 5 years, which would only be possible through intensive programs; not surprisingly all the students in that column are registered in such program (Figure 5.25). Most students started learning English between the ages of 6 and 10; although, as Figure 5.26 shows, there was a considerably high number of students who started when they were between 11 and 15 years old, especially in the intensive program. In general, it can be said that the students in that program started learning English a bit later than those in the extensive program. Similarly to what was reported for level 3, the level 5 students in the intensive groups seem to have had a more discontinuous experience studying the language than those in the extensive groups as Figure 5.27 and Figure 5.28 show. Regarding motivation, the students in the intensive level 5 program seem to be more motivated to learn English, as demonstrated by how much they like the language (Figure 5.29), and by the reason why they were studying English, which in most cases was a mixture of pleasure and necessity (Figure 5.30). There are more students in the extensive group in the categories '2' and '3' (in the five point scale from 'a little' to 'really a lot') when evaluating how much they like the language than in the intensive program; conversely, there are fewer students from the extensive program in the categories '4' and '5' than from the intensive (Figure 5.29). Moreover, most students in the extensive program were taking English because they needed it for their job or studies (free election credits), as observed in Figure 5.30.

Figure 5.25: Years Learning Eng Level 5

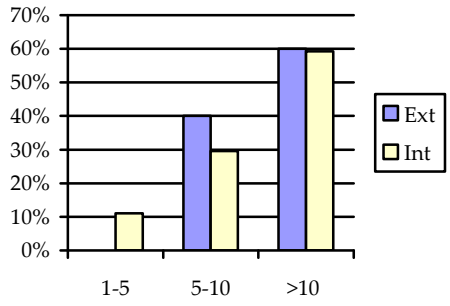


Figure 5.26: Age Started Eng Level 5

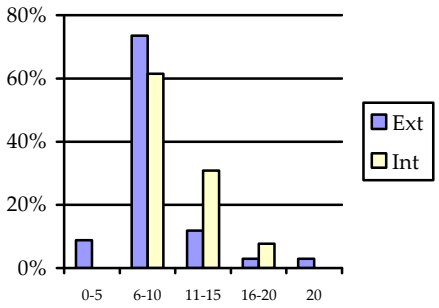


Figure 5.27: Continuity Eng Level 5

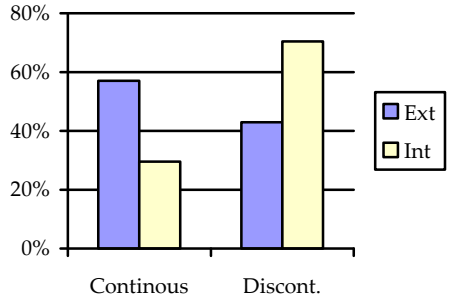


Figure 5.28: Previous Eng Cour. Level 5

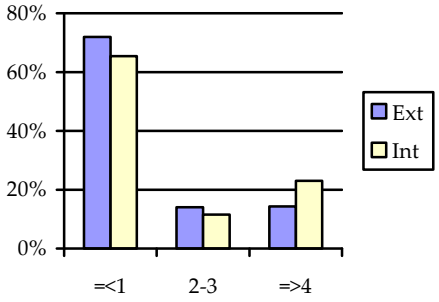


Figure 5.29: Like English Level 5

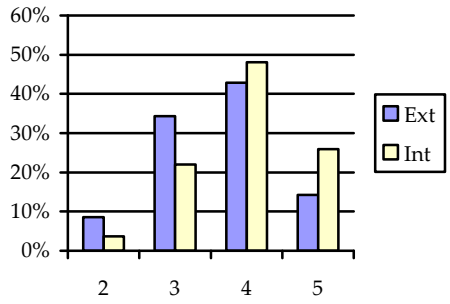
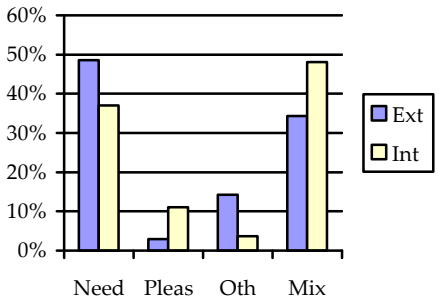


Figure 5.30: Reasons to Learn Eng Lev 5



All the students in extensive and intensive groups had had similar

experiences using the language in naturalistic contexts as illustrated in Figure 5.31 and Figure 5.32. Most students have practiced their English outside of Spain, and their stay in those countries was typically one month or less.

Figure 5.31: Eng-speaking Country L5

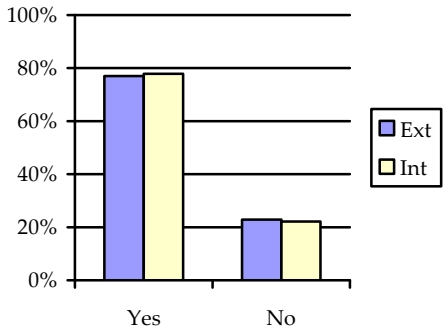
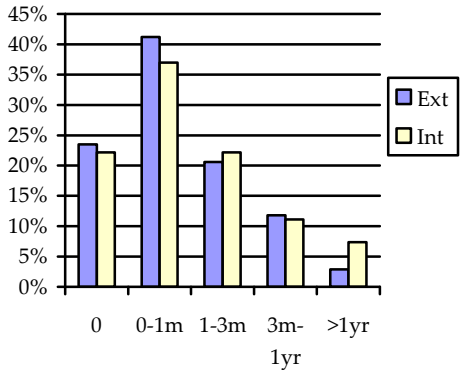


Figure 5.32: Time Eng-sp Count. Level 5



As was the case for level 3, the students in intensive courses attended their English classes more frequently than their peers in extensive groups (Figure 5.33), and they reported to have learned more in their course than students in the extensive program (Figure 5.34). With respect to the language areas practiced in class, students in extensive groups reported more speaking activities as well as a combination of other areas, whereas students in the intensive program mainly considered that in their classes they practiced many different skills (Figure 5.35), much more so than speaking or grammar and vocabulary. Curiously, more students in the intensive group reported spending more time doing grammar and

vocabulary than the students in the extensive group[34]. Figure 5.36 illustrates that, as expected, the students in the level 5 intensive program devoted more time to do homework each week than their peers in the extensive program.

Figure 5.33: Attendance Level 5

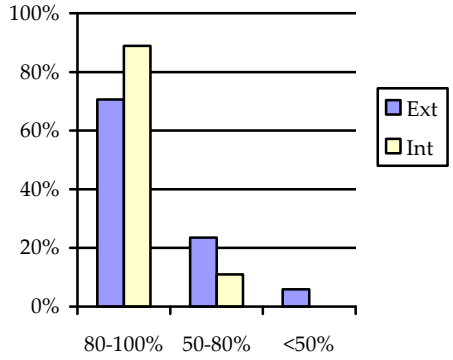


Figure 5.34: Learned this Year Level 5

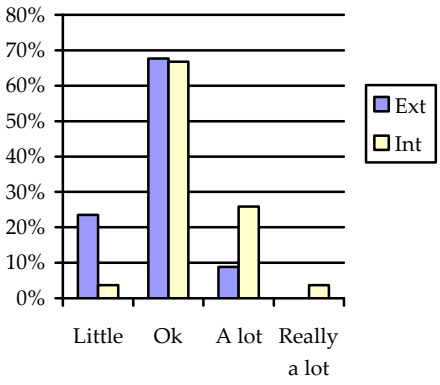


Figure 5.35: Classtime Lang Areas Lev 5

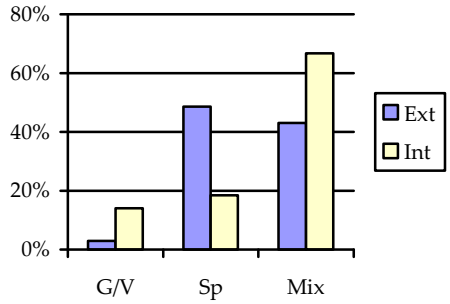
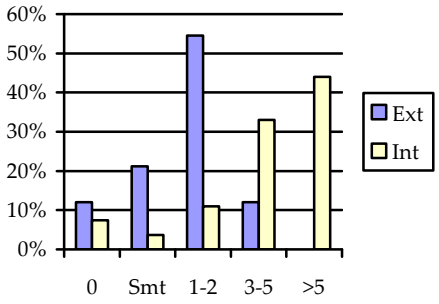


Figure 5.36: Hours/Wk Homewk Lev 5



In general it can be said that, with the exception of a very few areas (hours of homework a week, how much the students felt they had learned and in some

[34] This is due to the fact that in one intensive group vocabulary was given special emphasis, as reported by the students.

cases starting age and years learning English), most students' responses were quite similar across programs and also across levels.

With respect to the **teachers**, as can be seen in the column 'number of teachers' in Table 5.1, there is a varied number of instructors, almost one per group. There are four teachers who taught two different groups: teacher A taught level 3 extensive and level 3 semi-intensive, teacher C taught level 3 extensive and level 3 intensive, teacher I taught two level 5 extensive, and teacher F taught level 3 semi-intensive and level 5 intensive. As it was impossible to obtain groups in the three programs taught by the same instructor, it was decided to include as many instructors as possible, since such variation may neutralize the influence a specific teacher might have, which might be stronger than time distribution.

According to the information that the teachers provided in the questionnaire which they completed, most of the instructors were native English speakers (NES), as shown in Table 5.2; there were only two native Spanish speakers (NSS) and one native Catalan Speaker (NCC). Most of the teachers were between 41 and 50 years old, with a lot of experience teaching English (more than 15 years), although there were also a few instructors who were between 31-40 years old.

The methodology followed in the classes was similar. With respect to level 3, all groups followed the same textbook: *English File Intermediate* (OUP), which, according to the authors 'focuses on maximizing opportunities for students to speak'. The typical unit would start with a speaking or listening activity which is aimed at introducing the topic that will be dealt with. After such activity, more

language-focused exercises are performed, which aim at practicing vocabulary or grammar. As for the latter, the inductive approach is usually adopted, since the students are encouraged to do some practice exercises before actually learning explicitly the rules that apply in each case. Then, there is one reading through which the students can expand their knowledge of vocabulary and grammar, as well as see the vocabulary items or grammar points in context. In summary, it can be said that this book integrates quite well the four language skills and provides plenty of opportunity for the students to participate actively in the class through oral and written activities. Occasionally, songs, movies or TV shows are used in class in order to provide the students with natural input as well as to motivate them.

With respect to level 5, there was a wider variety in textbooks. The most popular ones were *Cutting Edge* (Longman), or *New Headway Advanced* (OUP). These two books are highly similar, since they both emphasize the four language skills, although the former can be said to focus more on vocabulary; more specifically, chunks and phrases. At the beginning of the units in both textbooks there are warm-up activities. These activities include pictures and speaking exercises, which target the students' previous knowledge of the topics mentioned in each unit, which tend to be current issues or subjects that are relevant for young adults. Then, there is usually one reading (which in *Headway* tends to be longer and more challenging than in *Cutting Edge*) in which the grammar and vocabulary points for each unit are introduced. Typically, some activities in each unit focus on writing skills. At the end of each of these books there is a grammar reference;

moreover, in the case of *Headway*, there is also a section on writing, while *Cutting Edge* includes extra grammar practice. At the advanced level, there is a strong emphasis on academic writing, formal speech (through classroom presentations) and fluency (in conversations in class). Realia are used in the classroom in the form of newspaper articles, movies, TV-shows, radio programs, etc.

Generally speaking, the methodology used in the three programs under study is quite similar, always considering that every teacher is unique in the implementation of a specific program. However, it must be mentioned that, since intensive programs are more demanding and require a lot of effort and concentration on the part of the student, more audiovisual activities are included in order to motivate the students and keep them more active.

A further group of participants was included in order to provide a baseline for comparison. Such group included 14 native English speakers, who were college students from the University of Illinois at Urbana Champaign (U.S.A.) following an intensive Spanish course at the University of Barcelona. These students' age, as well as fields of study, is comparable to the age and studies of the other participants included in this dissertation.

5.2. Procedure and instruments

The same data collection procedure was followed for the three different types of programs (extensive, semi-intensive, and intensive) and the two levels under analysis (intermediate and advanced), with a pretest/posttest design. The

students were not aware that they were going to be tested; therefore, their performance was not affected by last-minute study. The instruments used for collecting data can be divided in two types: written and oral. Section 5.2.1 will be concerned with the written instruments. First, section 5.2.1.1 will present the written instruments for the level 3 pretest, while section 5.2.1.2 will include the written instruments for the level 3 posttest. Next, the written instruments for the data collection at level 5 will be examined, first for the pretest (section 5.2.1.3) and then for the posttest (section 5.2.1.4). Section 5.2.2 will introduce the speaking instrument. Finally, section 5.2.3 will present the instruments which were used to collect data from the native English speakers.

5.2.1. Written instruments

5.2.1.1. Level 3 pretest

First of all, all the students took a written test at the beginning of the course (approximately 20 hours after the classes had started). This test was a short version of the practice exam used in the language school under research (with the same format as the final exam). Only a few exercises were chosen, since the testing was scheduled to last not much longer than one hour in order not to cause too much disruption to the usual classroom routine. The reasons why such test was chosen were mainly practical. First of all, the test gives the students a chance to prepare for the final exam; consequently, their participation in this study was expected to be

more enthusiastic. Additionally, it was thought that the teachers would be more in agreement with this research study, since they would not consider the time used for research as 'time wasted', since the practice being carried out was clearly related to the contents of the course. Moreover, since the activities in the test had been used as practice exam before, they were thought to be appropriate to test the contents taught in the levels under research. The written test included the following parts for the level 3 students (see Appendix A for the complete test):

- One listening comprehension exercise
- One sentence conversion exercise
- One open cloze activity
- One reading
- A 100-150-word composition

The students had approximately 60 minutes to complete these tasks. A schedule was presented to the students on the blackboard at the beginning of the test, to give them an idea about the time each of these activities should take. Obviously, the students worked at different paces, and some students finished the test within the proposed schedule, whereas others had to hand in their test without having completed all the activities included in it. The time allotted for each activity was the following:

- Listening comprehension exercise (10 minutes)
- Sentence conversion exercise (5-10 minutes)
- Open cloze activity (5-10 minutes)
- Reading (15-20 minutes)

- 100-150-word composition (15 minutes)

This test was administered either at the beginning or at the end of the class in the case of extensive classes. For semi-intensive and intensive groups, depending on the teachers' preference, sometimes it was administered right before or after the break. It was this researcher who administered the tests in all the groups, except in some of them where the teacher preferred administering the test herself, which was the case in one group in level 3.

For the **listening comprehension**, the students had to listen to a radio newsreader giving information about a competition. The students listened to this information twice and had to read several sentences related to the information presented, completing six blanks (worth 1 point each) with words the speaker used.

The **sentence conversion** exercise included five sentences (worth 1 point each) which the students had to paraphrase. The students were given the first words in the sentences. Several grammar points were tested in this activity: use of the passive, use of *enough*, linkers, modal verbs for requests and conditionals.

After this exercise, the **cloze activity** with the title 'A long journey' also tested the students' grammar knowledge: use of *never/ever*, *all/everything*, *another/other,* relative pronouns, verb tenses, modal verbs, *much/many*, and conjunctions. All these grammar points were a central part of the level 3 syllabus. The activity included 10 gaps, which required the use of one appropriate word each, and the students could obtain a maximum of 5 points (half a point for each correct word provided).

With respect to the **reading**, the students were given a text called 'British Seaside Resorts', which presented information on five tourist spots in Britain: Blackpool, Great Yarmouth, Morecambe, Brighton, and Sheringan. After reading the text, in Section A, the students had to decide which resort they would recommend to people who liked doing different activities: sunbathing, amusement parks, scenery, etc. Here the students had 10 items to complete and were given a maximum of 5 points. Section B provided five words for which the students had to find synonyms in the paragraphs indicated within the text. The maximum number of points in this section was also 5.

The last exercise in the written test was a 150-word **composition**, about the topic 'My best friend'. This topic was chosen because, first of all, it is a topic with which the students could easily identify; consequently, they would have many ideas to write and they would feel motivated. Additionally, it was thought to be simple enough for the intermediate students to be able to complete the task successfully.

5.2.1.2. Level 3 posttest

The posttest took place approximately 15 hours before the course finished, around 70 hours after the pretest. It was not possible to test the students later for several reasons. First, and most importantly, the last few classes are typically left for exam preparation and addressing the doubts the students might have before taking the exam. Second, some teachers may need these classes in order to finish

the syllabus. It is very common for teachers to be stressed towards the end of the course due to the lack of time to deal with certain points in the program.

The same activities were used in the posttest as in the pretest for the listening, sentence conversion, cloze and reading, in order to measure progress in a reliable way, since there was not enough time to develop other activities which measured the same abilities in the pretest and in the posttest. It is true that the testing effect cannot be ignored, and the fact that the students repeated the same activities after a few weeks (in the case of intensive students) or a few months (in the case of students in extensive programs) might have had an effect on the results, as some studies have demonstrated in cases of task repetition (Bygate, 2001; Yuan & Ellis, 2003). However, this was the most appropriate design that could be developed when planning this project; moreover, as Larsen-Freeman (2006) reports, having different tasks in the pretest and in the posttest can also affect performance. Besides, more fine-grained changes can be captured more easily if the same test is used several times than if a new test is created.

For the composition, the topic was changed to 'Someone I admire'. The reason why the topic of the composition was changed, while the other exercises in the written exam were not, was due to the fact that the students were expected to remember more clearly the subject of a composition from one time to another than the testing of a specific grammar point. Repeating the same composition was probably not motivating for the students, and that is why the topic was changed. At the same time, the differences between the two topics suggested in this study were not considerable; both times the students had to talk about one person (in

some cases it was their best friend who they admired), describe them and present reasons why that person was their best friend (pretest) or their 'idol' (posttest).

The procedure followed in the posttest testing was the same, except that 15 more minutes were provided for the completion of a questionnaire in Spanish, to make sure that the students understood the questions, (see Appendix C), which aimed at gathering some information about the students' background (age, studies, etc.), English knowledge (starting age, years of study, experiences abroad, reasons for studying it, etc.), knowledge of other languages, opinion about the course they attended, previous experiences in intensive courses, etc. This information was summarized in section 5.1. The teachers were also given a questionnaire, which aimed at obtaining some background information about themselves, but mainly about the course they had taught, and about other intensive courses in which they had provided instruction (see Appendix D).

5.2.1.3. Level 5 pretest

The test that the advanced students completed, which took approximately 60 minutes, was quite similar to the one taken by the level 3 students. The activities included and the time suggested for each activity was the following (see Appendix B):

- One listening comprehension exercise (15-20 minutes)
- One sentence conversion exercise (5-10 minutes)
- One open cloze activity (10 minutes)

- A 100-150-word composition (20 minutes)

This test was also administered either at the beginning or at the end of the class in the case of extensive classes, or before or after the break in the intensive groups. The researcher administered these tests in all the groups, except in one, in which the teacher asked to do it himself.

For the **listening** comprehension exercise the students had to listen to a man explaining the different means of transport which can be used to travel around Turkey and the advantages and disadvantages of traveling by train or by bus. After listening to the speaker twice the students had to complete an activity which consisted of nine sentences related to the information previously heard, but, as opposed to the equivalent exercise in the level 3 test, the sentences included here did not reproduce the exact same words used by the speaker; therefore, it was more challenging, since it required a good understanding of the information presented and not simply the recognition of some words. Each of the nine sentences had one gap worth 1 point, and correspondingly the whole exercise was worth a maximum of 9 points.

Then, in the **sentence conversion** exercise the students had to write equivalent sentences to the ones they were provided, once they were given the first words in the sentences. This exercise included five sentences (each worth 1 point) which tested the following grammar issues: verb complementation, *I wish*, passive voice, use of linking words and conditionals.

The **cloze activity**, 'The ark in the dock', was also worth 5 points and had 10 gaps which had to be filled with one word. The grammar and vocabulary items

tested included modal verbs, verb tenses, collocations, relative clauses and linking words.

The last activity the students performed was a 150-word **composition** about the topic 'My best friend'. The reasons why this topic was used were the same as the ones presented in the corresponding section for level 3 (section 5.2.1.1). The same topic was used for the advanced and the intermediate learners since it was considered appropriate for any proficiency level. Moreover, including the same task at levels 3 and 5 could facilitate comparisons between the two groups of learners.

5.2.1.4. Level 5 posttest

The level 5 posttest also took place towards the end of the course, approximately 70 hours after the pretest and at a time when the students had received a total of 80-100 hours of instruction since the beginning of their course. The reasons why the test could not be scheduled later were the same as those presented for the intermediate students. Apart from that, in level 5 the students have to do presentations in front of the class, and a lot of those presentations take place towards the end of the course.

As was the case for level 3, the same materials were used in the posttest for this level too, with the exception of the topic for the composition, which was changed to 'Someone I admire'. Similarly, for the posttest the students were left 15 more minutes than in the pretest in order to complete the background

questionnaire, which was the same as the one provided for the level 3 students.

5.2.2. Speaking instrument (levels 3 and 5)

Once the written data was collected (at both points of the data collection, pre and posttest) a subset of the students in each group took an oral test. Usually the speaking test was performed in the class session following the written test (for extensive and semi-intensive groups), or on the same day as the written test (in the case of most intensive groups).

Table 5.3: Students Who Took the Speaking Test

	Students per group	Students per program
3J (Ext)	6	21
3V (Ext)	7	
3L (Ext)	4	
3X (Ext)	4	
3Semi1	0[35]	11
3Semi2	6	
3Semi3	5	
3IntA	7	21
3IntB	7	
3IntC	7	
5G (Ext)	7	23
5H (Ext)	7	
5I (Ext)	3	
5N (Ext)	6	
5IntA	8	22
5IntB	7	
5IntC	7	
TOTAL	**98**	**98**

[35] For administration problems it was impossible to collect oral data from the 3Semi1 group.

Table 5.3 illustrates the number of students who did the oral test in the pre and posttest and the total number of students per program type and level, together with the total number of participants who underwent the oral testing.

The students were chosen randomly for this activity (except for the intensive groups, where virtually all the students took the oral test). The researcher called the students one by one and asked them to leave the class (where the instructor continued teaching), and follow her into a separate area where they performed the task. The students were recorded while they told a narrative on the basis of a series of pictures called 'The dog's story'. This test was used originally in the BAF project (see Muñoz, 2006), and, since then, it has been used in a variety of projects including learners with different L1's and L2's. The story (see Appendix G) shows how two children are preparing some food with their mom in order to go out for a picnic. While the mom is explaining to the children how to get to the picnic area with the help of a map, their puppy hides into a basket, where they have put their sandwiches, and eats all their food. Consequently, when the children are in the country and want to eat their sandwiches, they are surprised because their dog appears suddenly and they realize that there is no food left. This story is always motivating since all the students, no matter what their proficiency level is, can say something about it: the students with a lower proficiency level can simply describe the pictures, whereas those learners with a more advanced command of the L2 can narrate the story.

This researcher was in charge of the collection of the oral data for approximately 75% of all the students, the remaining 25% was completed by four

research assistants, who followed the same procedure for the data collection. The students usually took around two minutes to narrate the whole story. Once a student finished with this task, she/he went back to the classroom and called another student, who did the same as the previous one.

5.2.3. Instruments for native English speakers

Since the instruments used in this dissertation to collect samples of learners' written and oral production at levels 3 and 5 included tasks which are more open, and more holistic in their evaluation than those used to evaluate the students' listening, grammar, vocabulary and reading skills, it was decided that having a baseline for comparison for those tasks from native English speakers would be appropriate. Native English speakers were asked to write a 150-word composition on the topic 'My favorite city' or 'Someone I admire'. The students spent around 10 minutes to complete this task. Then, all the native speakers performed the oral narrative 'The dog's story' with a research assistant or this researcher herself.

5.3. Measures

5.3.1. Level 3 listening, grammar, and reading sections

The written test that the level 3 students took was worth a maximum of 26 points (see Appendix A for the actual test). First, the listening exercise was scored

with one point for each of the six gaps that the students had to fill in. No point was given in the case of spelling errors which would alter the pronunciation of the appropriate word (e.g. *crim* intead of *crime*), and half a point was granted when the word that the students provided minimally altered the grammar of the whole sentence it belonged (*future* instead of *the future* in number 4).

With respect to the sentence conversion activity, one point was given for each correct paraphrase of the five sentences provided, and half points were supplied in the case of sentences which were grammatical and highly similar in meaning to the original, yet some meaning nuance was omitted (e.g. *Because of the weather they couldn't go skiing* from *The weather was foggy so they couldn't go skiing*)

The cloze exercise was worth 5 points, half a point for each of the ten blanks the students had to complete. There was only one possibility for each blank, except for number 3, where both *one* and *flight* were acceptable and worth 0.5 points.

The reading activity measured two different skills: comprehension of a text (Section A) and knowledge of vocabulary (Section B) (see 5.2.1.1 for details on this activity). Section A included ten different gaps (scored with 0.5 each), where the students had to include the name of tourist towns which were more convenient for a person according to the preferences given. There was only one possibility for each of the blanks. Section B was also worth a maximum of 5 points. Only one synonym could be provided in order for the students to obtain the 0.5 points that each word was worth.

The tests were all scored two times after the acceptability criteria were designed in order to eliminate inconsistencies. The intrarater reliability, which was

calculated using the percentage of agreement, score was 97%.

5.3.2. Level 5 listening and grammar sections

As a whole, the exercises assessing listening and grammar skills at level 5 were worth a maximum of 19 points (see Appendix B for the actual test). Each of the nine blanks in the listening comprehension exercise was worth one point. Since, contrarily to the equivalent exercise in the level 3 exam, the students did not simply have to write words used by the speaker, there were several alternatives which were positively considered for most blanks. In detail, the words which were accepted, partly accepted or rejected for each of the blanks can be consulted in

Appendix E.

The criteria for scoring the sentence conversion exercise were similar to the ones presented in the previous section for level 3. One point was given for each correct transformation and half points were provided in cases where the sentence created by the students was grammatical and highly similar in meaning, but lacked some aspect included in the original sentence. The total number of points for this section was 5.

Finally, the cloze exercise was worth a maximum of 5 points (half a point for each of the ten blanks). There was only one possibility for most of the blanks; however, in some cases different alternatives were accepted. For instance, in number 2) *some, nearly, almost, about, around, approximately* or *over* were all accepted, in 8) *might, may, could, would* and in 9) *consequence, result* (see Part 3, Appendix B).

As for the level 3 test, these activities were all scored two times after the acceptability criteria were designed in order to eliminate inconsistencies. The intrarater reliability, which was calculated using the percentage of agreement, score was 96%.

5.3.3. Writing measures (level 3 and 5)

Measuring writing proficiency is usually a more complicated enterprise than scoring grammar or listening tests. Different measures have been used which have intended to account for writing proficiency or writing development across time. Most of these measures relate to three main areas which have been examined

in terms of written production: fluency, accuracy and complexity (both lexical and grammatical). This section will provide general explanations of the measures and the reasons why they were adopted. Measuring written production can be holistic by means of scales or impressionistic grading (which most ESL/EFL teachers do when correcting their students' writing samples) or analytic, through objective, quantifiable measures. Holistic grading can be claimed to be highly subjective, since giving a global mark for a composition does not follow a set of objective criteria, the same as allocating a specific sample within a certain range in a particular scale. At the same time, holistic grading can consider certain aspects that are certainly missed when more objective measuring is selected, such as the "sense of fluency, maturity of topic, depth of treatment, honesty and genuineness of opinion, ability to take on a new perspective and reflectiveness" (Casanave, 1994: 181). Some examples of holistic scales include the American Council for the Teaching of Foreign Languages (ACTFL) Proficiency Guidelines and the Common European Framework Reference for Languages (Council of Europe, 2001). Additionally, some projects have developed their own holistic measuring taking into account their samples, such as the "Two-Way Immersion Narrative Writing Assessment Rubric" developed at the Center for Applied Linguistics and used in order to account for writing development in Spanish and English for native Spanish and English speakers in their L1 and L2 (Howard, 2003; Serrano & Howard, 2007). As mentioned in Homburg (1984) one of the main problems with holistic measures is reliability (both inter and intra-rater).

Analytic measures, on the other hand, are generally more objective and

easily generalizable across contexts, since they usually involve counting specific units chosen to describe certain aspects of writing proficiency. However, it is precisely that pre-supposed relationship between proficiency and the suggested measures which has been most criticized regarding analytic measuring, since, for instance, the fact that the total number of words is an index of fluency is debatable, despite the fact that every rater would agree on the result of such measure (composition A has 300 words). As a consequence, there is also, to a certain extent, some 'intuition' involved in analytic measures, yet such intuition is not so much in the scores, as in the operationalization of the writing areas under investigation (typically fluency, complexity and accuracy). Some studies which have considered a high number of analytic measures to analyze writing samples from school EFL learners include Celaya, Pérez-Vidal, & Torras (2001), Navés (2006); Pérez-Vidal, Torras, & Celaya (2000), or Torras et al. (2006).

Despite the different perspectives and criteria adopted for holistic or analytic measuring, the choice of one type of measures or the other may not be so influential in the results, as some studies have shown that equivalent results are found when comparing the scores using holistic or more objective measures (Homburg, 1984).

In this particular research, analytic measures have been chosen for a variety of reasons. First, they are considered highly reliable, since there have been a number of research studies that have shown that the operationalization of fluency, accuracy and complexity in the analytic measures which have been suggested is consistent. Next, both inter and intra-rater reliability are believed to be more easily

reached using objective measures. Additionally, as some researchers have claimed, holistic scales are not specific enough as to discriminate between members of a homogeneous population (Polio, Fleck, & Leder, 1998), and by implication, development over a short period of time (which is the case of the participants in this research study).

Wolfe-Quintero, Inagaki, & Kim (1998) made a thorough review of the most commonly used analytic measures for writing development, explaining how the results obtained using such measures correlated with levels of proficiency. One of the main purposes of this review was to investigate which measures accounted better for development in writing. The researchers considered 39 second/foreign language studies which analyzed writing development (not writing proficiency per se) in communicative situations; therefore, samples such as narrative or persuasive essays, picture or film descriptions, journal writings, etc. were included. This review will be used as a guide in order to discuss the measures adopted for this particular research study.

In order to choose among fluency, accuracy or complexity measures, the first decision that should be adopted is whether frequency, ratio or index scores are more appropriate for the research study being performed. Frequency calculations include the total number of units (words, T-units, sentences, errors, etc.) in a sample. Ratio measures, on the other hand, relate units to other reference units (words *per sentence*, sentences *per T-unit*, errors *per clause*, etc.). Finally, index measures are based on formulas that also relate two units, yet in a more complex way than ratio measures; for example, the coordination index as described in

Wolfe-Quintero et al. (1998: 95) is calculated "by dividing the total number of independent clause coordinations by the total number of combined clauses, which are the number of clauses minus the number of sentences".

Frequency calculations are quite related to the time which was allotted for task completion; as a consequence, such calculations were not adopted in this study for the following reason. In this research the participants did not use the same amount of time to complete the writing task, since the students were given the topic of the composition together with the rest of the written tasks; consequently, they could decide the amount of time they wanted to devote to each of the exercises (despite the fact that a 'suggested schedule' was presented on the blackboard for all the students, as explained in sections 5.2.1.1 and 5.2.1.3). Indices have been less commonly used and are highly more complex than ratios. Therefore, ratio calculations were considered to be the most adequate for this study, thus all the measures adopted to analyze progress in writing ability in terms of fluency, accuracy and complexity will be ratios. Such ratios will always take the T-unit as the production unit.

The T-unit was developed by Hunt (1965), as an alternative to the sentence, since the latter is subject to the learner's knowledge and command of the punctuation system of a specific language. The T-unit is defined in Hunt (1965: 20) as "one main clause with all subordinate clauses attached to it". This definition is quite general and does not specify whether coordinated clauses within subordinate clauses should be included in the same T-unit (e.g. *When they opened the door and saw their neighbor they were scared* or *They realized that Tom had been there and that he*

had eaten all the cookies). In the present study, coordinated subordinate clauses will be counted within the same T-unit; therefore, the two examples previously presented would be considered as one T-unit each. The reason for such decision is that the coordinated subordinate sentences are part of the same syntactic unit (complement in the first case and direct object in the second), which is dependent of the main clause, hence it was considered more appropriate to analyze the two clauses within the same unit. Additionally, in the definition of T-unit there is no specific reference to whether punctuation should be taken into account or not. In this study, the T-unit will be unaffected by the students' punctuation, thus *When she visited him in Chicago. He was very happy* will be considered one T-unit, even if the two clauses are separated by a period.

In the next sections the measures adopted to analyze fluency, complexity and accuracy will be presented. After providing a brief introduction to the different types of evaluation that have been used, the measures adopted for this particular study will be discussed.

5.3.3.1. Fluency measures

Fluency has been defined through different concepts related to speed of language production, automaticity (Ellis, 1996; Hulstijn, 2001; Schmidt, 1992), ease of retrieval of language items, and length of output (Lennon, 1990), or even a mixture of different qualities related to speed, coherence, appropriateness and creativity (Fillmore, 1979). According to Wolfe-Quintero et al. (1998: 14), "fluency

means that more words and more structures are accessed in a limited time, whereas a lack of fluency means that only a few words or structures are accessed". This definition, which only addresses the issues of rate and length, disregarding others as appropriateness or creativity, will be adopted for this research. Some measures of fluency which have been used in different studies have been the following: words per minute, words per clause, words per sentence, words per T-unit, words per error-free T-unit, words per error-free clause, words in complex nominals per T-unit, and words in complex nominals per clause, with **words per T-unit** (which is the measure chosen for this study) and words per error-free T-unit as the most commonly used ratios (Wolfe-Quintero et al., 1998). One reason to adopt the former for this study is that several studies have considered this measure as one of the best to describe development in second language writing (Larsen-Freeman, 2006; Larsen-Freeman & Strom, 1977; Wolfe-Quintero et al., 1998). Moreover, if the ratio words per error-free T-unit had been selected, some problems would have appeared at the time of scoring samples which lacked error-free T-units (which were not common within the data under examination, but certainly existed); therefore, if the measure of fluency was related to accurate units, those samples would have to be scored with '0 fluency', disregarding the rate and amount of the language produced by a certain student. The aim of choosing a fluency measure was to analyze fluency per se, and not to relate it with other aspects of written production, such as accuracy.

It must be indicated that some researchers have claimed that the ratio words per T-unit (W/T) measures grammar complexity more than fluency.

Nevertheless, I agree with Cooper (1976) and Wolfe-Quintero et al. (1998) that longer does not necessarily mean more complex, since longer T-units can incorporate a higher number of words and phrases, yet such words do not have to be included necessarily in complex grammatical clauses (even if in many cases some measures of grammar complexity, such as clauses per T-unit, clearly correlate with W/T, as will be seen in the results of this research study). O'Donnell (1976: 33) argued that "the T-unit can be lengthened in a variety of ways, some of which require a great deal more linguistic maturity than others". Some evidence for the fact that longer T-units do not need to include more complex clauses is found in Casanave (1994), who observed that many of her students after some hours of instruction produced longer T-units, but they were less complex and more accurate.

5.3.3.2. Complexity measures

Complexity in writing can be analyzed in terms of grammar or in terms of vocabulary. In this section, grammar complexity measures will be examined first, and then measures of lexical richness will be presented.

Grammatical complexity refers to how elaborate the syntactic patterns used by the learners are. At higher levels of proficiency, once simpler patterns have been automatized, learners can devote more resources to the restructuring of higher syntactic levels (Ellis, 1996; McLaughlin, 1990); as a consequence, grammar in the more advanced stages of second language acquisition should be more complex than at earlier stages. Following Hunt (1965) the stages in writing development

with respect to grammatical complexity are the following:

1. fragments
2. main clauses
3. coordinate clauses
4. adverbial clauses
5. adjective and nominal clauses
6. adjectival, adverbial, and nominal verb phrases

According to these stages, more cases of simple or coordinate clauses should be expected in the case of lower proficiency learners, and more subordination when examining advanced learners' writings. Wolfe-Quintero et al. (1998) present some grammatical complexity ratios with respect to whether they are general complexity measures (clauses per T-unit, which is the measure adopted in the present study to analyze development in writing in grammatical complexity, or clauses per sentence), whether they measure the relationship between dependent and independent clauses (e.g., dependent clauses per sentence or T-unit), or the relationship between coordination and independent clauses (e.g., coordinate phrases per T-unit).

In order to analyze development in syntactic complexity the **clauses per T-unit (C/T)** ratio has been adopted in this study, and within the term 'clauses', both finite and non-finite clauses are considered. Wolfe-Quintero et al. (1998: 86) claim that the majority of the studies reviewed by them "do support the usefulness of the clauses per T-units measure", despite the fact that some studies did not find correlations between proficiency and the C/T ratio. In this particular research study

chunks and fragments which did not include verbs (finite or non-finite) were not considered as clauses; they were classified as [chunks] or [fragments] and excluded from the analysis of syntactic complexity. Since the data was gathered from mature learners with some knowledge of the English language and experience writing and speaking, there were very few instances of fragments or chunks in the students' production.

Lexical richness has been analyzed in terms of lexical variation with measures like the Type/Token ratio, lexical density (e.g., lexical words per words) or lexical sophistication (as sophisticated verb types per verbs). In this study only lexical density will be examined, using the **Guiraud's Index of Lexical Richness**: word types divided by the square root of the word tokens (WTypes / √WTokens)[36]. Some studies have shown this measure to be one of the most adequate to analyze lexical richness in L2 learners' productions (Vermeer, 2000). In her review of the most commonly used measures of lexical richness in spontaneous speech data, Vermeer (2000) concludes that the Guiraud's Index is one of most adequate measures, while the traditionally used Type/Token ratio was claimed to lack validity and reliability. Nevertheless, despite many positive reviews of this index, it has also undergone some criticism (Daller, Van Hout, & Treffers-Daller, 2003) due to the fact that it only analyzes vocabulary quantitatively and a mixed approach including quantitative and qualitative measures has been claimed to be more effective.

[36] In order to examine how the writing samples were transcribed before the Guiraud's Index was calculated, see section 5.3.4.4.

Most criticism of the Type/Token ratio relates to the fact that it is a measure which is sensitive to text length; as a consequence, the ratio decreases for longer writing samples, since it is in those samples where there is a higher possibility for the repetition of words. Apart from Guiraud's Index, other alternatives have been suggested to counterbalance the effect of text length. The ratio word types divided by the square root of two times the total number or word tokens (WTypes / √2WTokens) developed by Carroll (1967) has been used in a variety of studies (Larsen-Freeman, 2006). Another measure to assess lexical richness is D (Malvern et al., 2004), which has been claimed to control for text length. It has been used in a variety of research studies related to foreign language acquisition (Miralpeix, 2006). The Guiraud's Index, as well as the D, makes use of all the vocabulary in the text.

On the other hand, there have been some researchers who have reduced the length of all their samples to the length of the shortest one, assuming that, when the Type/Token ratio is calculated for one particular learner's reduced sample, one can assume that such ratio would still hold for longer samples of the same learner (Arnaud, 1992). The problem of adopting this alternative for the present study is, first of all, that the number of words the students were supposed to write was not especially high (maximum 150); therefore, reducing the number of words from such a set would leave us with small samples. Additionally, since the students had a limited time to do the grammar and reading exercises and the composition, some students did not have sufficient time to write the minimum of 100 words; therefore, if the shortest composition were to be taken as the model to reduce the

other samples, the result would be extremely short writing samples for analysis. Therefore, the Guiraud's Index was considered the best measure, since it allows the researcher to keep the length of the compositions as the students designed them, yet text length is not an intervening variable when examining lexical richness.

5.3.3.3. Accuracy measures

According to Wolfe-Quintero et al. (1998: 33), accuracy is "the ability to be free from errors while using language to communicate in either writing or speech". Usually, second language errors are deviations from target-like utterances; therefore, the native speaker is the model for accurate performance, which some researchers have criticized (Thomas, 1994). According to Ellis (1996) or McLaughlin (1990) accuracy is related to language automatization, since the more automatic language production is the more fluent and accurate it becomes. Similarly, Newell & Rosenbloom (1981) and Anderson (1993) also suggest that, according to the power law of practice, performance in any skill is faster and more accurate as such skill is automatized due to extensive practice.

According to the claims presented before, it seems reasonable to expect that, as learners make progress in their second/foreign language knowledge, their language production should be more accurate. Nevertheless, such expectation does not always correspond to the reality. As learners' language becomes more complex and fluent (both in written or oral production), there are higher opportunities for errors. Indeed, some studies have found that whereas beginners were more likely

to make errors than other groups of learners, advanced students have been found to make more errors than intermediate learners both in written and oral production (Foster & Skehan, 1996; Larsen-Freeman, 1983; MacKay, 1982). It can be the case, though, that the learners examined in the previously mentioned research studies had an advanced knowledge of the L2; nevertheless, such knowledge was not automatized yet. It can be said that it is not until automatization is complete that the L2 learner's production can be highly or completely accurate. Other studies, however, have found that increasing the complexity of the writing tasks produced more complex language without it having any negative impact on accuracy (Robinson, 2003b; Shuqiang, 1989).

A wide range of ratios has been suggested to measure accuracy in second language learning, depending on whether the researcher's interest is to analyze errors in general (error-free T-units per T-unit, error-free T-units per sentence, error-free sentences per sentence, error-free clauses per clause, errors per T-unit, etc.), types of errors according to their degree (first-degree errors per T-unit, second-degree errors per T-unit, etc.), types of errors according to the language area (syntactic errors per clause, morphological errors per clause, lexical errors per lexical word, etc.), or whether the focus was more on correctness (correct number of connectors per total number of words, correct definite articles divided by the total number of contexts either required or supplied, etc.)

Out of all these measures the most commonly used have been **error-free T-units per T-unit** (henceforth EFTU/T) and **errors per T-unit** (Err/T), which will be the ones used to examine accuracy in this study. It should be born in mind that the

number of errors does not necessarily decrease as proficiency level increases (as was mentioned above), and also that these measures have been reported not to capture short-term change within intact classes (Wolfe-Quintero et al., 1998).

Whereas for fluency and complexity only one measure was chosen, in the case of accuracy it was considered more appropriate to select these two measures, since they provide different types of information. The EFTU/T can present a general picture of accuracy taking into account the T-units; however, this measure does not discriminate between T-units which have one or many errors; that is the reason why it was thought appropriate to analyze errors as units using the Err/T measure.

Accuracy measures, whether they are global, like the ones proposed for this particular study, or more local measures, focusing on errors on specific language areas (articles, past tense, etc.), are highly problematic since they are more prone to subjective interpretations than fluency or complexity measures, regarding what counts as an error and the 'seriousness' of a specific error. Wigglesworth & Foster (2007) highlight these problems with measuring accuracy and propose a measure which aims to be more reliable and includes three levels of errors, assigning 1 point to error-free clauses, then 0.8 points to level 1 errors (which the authors consider 'minor errors'), 0.5 points to level 2 errors (errors which are more serious but do not hinder communication), and 0.1 points to level 3 errors (when learner's intended meaning is hard or impossible to recover). Although I agree with Wigglesworth & Foster (2007) that more reliable and valid measures of accuracy are necessary, it is also true that the levels suggested by these authors face the

same problems researchers faced when counting errors, in the sense that there are no clear, objective criteria as to what a level 1, 2, or 3 error is.

Summarizing, the following measures have been used in this study in order to analyze the students' writing samples in terms of both the progress made by all the students from the beginning until the end of their respective EFL courses, and in terms of how intensive English instruction affects such progress:

- **Fluency**: words per T-unit (W/T) (number of words divided by the number of T-units)
- **Complexity**:
    - ***Syntactic***: clauses per T-unit (C/T) (number of clauses divided by the total number of T-units)
    - ***Lexical***: Guiraud's Index (number of word types divided by the square root of the number of word tokens)
- **Accuracy**:
    - Percentage of error-free T-units (EFTU/T) (number of error-free T-units divided by total number of T-units x 100)
    - Errors per T-unit (Err/T) (number of errors divided by the total number of T-units)

All of these measures have been widely used in the second language writing literature research, and usually many of them have been included in the list of the best potential measures for second language writing development (Larsen-Freeman, 1983; 2006; Larsen-Freeman & Strom, 1977; Wolfe-Quintero et al. 1998).

5.3.3.4. Coding and statistical analyses

The writing samples which were collected from all the students included 150-word compositions on the topic 'My best friend' (pretest) and 'Someone I admire' (posttest). In order to analyze those writings, they were initially transcribed using Microsoft Word by this researcher. Then, the CHILDES program (MacWhinney, 2000) was used in order to analyze the compositions[37]. First, the writings were copied and coded in CHA format[38]. This researcher was in charge of the whole process; additionally a research assistant coded a random 10% of all the samples. Interrater agreement calculated by means of percentage agreement reached 93%. Intrarater reliability reached 96%.

After the coding process was finished, the CLAN program was used in order to analyze the frequencies of each of the units considered for the writing samples (total number of T-units, clauses, error-free T-units and errors). Then, the Statistical Package for the Social Sciences (SPSS) was used in order to analyze the data. First, the frequencies obtained using the CLAN program were included into an SPSS matrix, and the formulas were calculated for the specific measures (W/T,

[37] See Appendix F for an example of a transcription.

[38] All the written material was transcribed as produced by the students. Spelling mistakes were not considered errors, since it was not one of the objectives of the analyses to examine spelling. Moreover, spelling mistakes were corrected, because a word appearing once mispelled and another time properly spelled would count as two different tokens in the frequency analysis, which would lead to inaccuracies in lexical measures. Words in the students' L1 were kept (and considered as errors), yet they were marked and excluded from the analyses of types and tokens.

C/T, Guiraud's Index, EFTU/T, Err/T). Afterwards, ANOVAs and other statistical analyses, including parametric and non-parametric tests, were performed and the final results were obtained.

5.3.4. Oral production measures (level 3 and 5)

The students' speech production was analyzed in terms of fluency, complexity (grammatical and lexical), and accuracy. All the measures used in each of these areas are the same as those used for the writing samples, except in the case of fluency. Also, ratios will be adopted and not frequency measures or indexes.

In spoken language, several production units have been normally used when establishing ratios. Apart from the T-unit, other measures have been proposed, since the analysis of oral data presents difficulties that are not usually encountered in the written mode, such as false starts, repetitions, self-corrections, or interaction features, as confirmation checks. That is one of the main reasons why the AS-unit has been suggested, which is defined by its creators as "a single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either" (Foster, Tonkyn, & Wigglesworth, 2000: 365). In their study these researchers mention several problems that can appear when measuring with T-units, and then 'solve' those problems with the measure they developed.

Other measures of speech production have been the utterance or the C-unit. The utterance has been defined as a sequence of speech "produced under a single

intonation contour bounded by pauses" (Sato, 1988: 375). Ortega (1999) used the ratio number of words per utterance. Another reference measure, apart from the T-unit and the utterance is the C-unit:

> A c-unit is defined as each independent utterance providing referential or pragmatic meaning. Thus, a c-unit may be made up of one simple independent finite clause or else an independent finite clause plus one or more dependent finite or nonfinite clauses. (Foster & Skehan, 1996: 31)

For this particular study, the T-unit was preferred for several reasons. First of all, it was the production unit adopted for writing performance, thus parallelisms can be established between written and oral production. Second, some of the problems that have appeared when trying to use T-units in spoken data have not been found in the samples considered for this research, since the data included narratives, where no interaction occurred. Moreover, since the students' level was at least intermediate, they generally formulated whole T-units[39].

As was claimed in section 5.3.3.3, there are usually trade-offs between fluency and accuracy (Yuan & Ellis, 2003), or complexity and accuracy (Mehnert, 1998; Skehan & Foster, 1997), especially at lower proficiency levels. This means that non-proficient learners, because of the limits of their processing capacity and their need to make decisions about where they allocate their attention, will either focus on meaning, (or 'message conveyance') thus producing more fluent and lexically richer speech; or on form, using more grammatically accurate sentences (Yuan & Ellis, 2003). VanPatten (1990) claims that non-proficient L2 learners either pay attention to form or meaning, and when they have to decide where to allocate their

[39] The few chunks which were produced (fragments with no verbs) were not considered for syntactic complexity.

attentional resources, learners tend to focus on meaning with the corresponding decreasing accuracy on forms. Robinson's view, on the other hand, (Robinson, 2001; 2003a; 2003b) is that attention can draw on multiple resources (following Wickens, 1989); consequently, increasing the cognitive demands of a task can also increase syntactic complexity in oral production without it having negative effects on accuracy.

5.3.4.1. Fluency measures

As reported in Griffiths (1991) the most widely used variables of fluency are the speech rate (words per minute being the most common measure), articulation rate (syllables divided by articulation time, which is the total time minus the pauses), and other measures related to silent pauses, such as their frequency, duration or distribution. In order to measure pauses, more sophisticated instruments than a stop-watch are required if precise results are the objective; moreover, digital recordings are highly desirable when more refined measurement is necessary. Since the recordings for this study were made using tapes, it was considered that the most accurate measure of fluency would be **syllables per minute**. Moreover, this measure has been used in a variety of studies of L2 oral production (Gilabert, 2005).

When using the oral fluency ratio **syllables per minute (syll/min)**, the researcher needs to take decisions concerning the syllables which should be

included in the count, and whether false starts, repetitions, self-corrections, or words in the student's L1 should be considered for the total number of syllables or not. Two speech rates have been proposed: Rate A, which includes all the syllables uttered by the student; or Rate B, which does not consider false starts, repetitions, self-corrections, or words in the L1 (Gilabert, 2005; Yuan & Ellis, 2003). Both of these measures were initially considered in this study; nevertheless, as the differences between groups or across time using the two measures were similar, only the Rate B will be reported, first for the sake of simplicity, and second, because it was assumed that such rate defined better the students' L2 fluency, since it eliminates non-meaningful syllables.

5.3.4.2. Complexity measures

Previous studies of oral production have investigated grammar complexity analyzing the number of clauses (or sentence nodes, which is usually considered synonymous) according to the production unit. For example, Foster & Skehan (1996) counted the clauses per C-unit, whereas Mehnert (1998) or Gilabert (2005) used the sentence-nodes per T-unit measure. In this particular study the **clauses per T-unit (C/T)** measure, which considers both finite and non-finite clauses, was adopted to analyze complexity in oral production for several reasons. First of all, it was the measure used for the written samples; consequently, it was believed that having parallel units would produce more comparable results. Moreover, grammatical complexity in oral production has usually been measured by

analyzing clauses, sentence nodes, or subordinate clauses according to the reference unit adopted.

For lexical richness, the **Guiraud's Index of Lexical Richness** (WTypes / √WTokens) was considered the most appropriate measure for the reasons presented in section 5.3.3.2. Several studies have used this measure in order to examine lexical richness in L2 oral production (Gilabert, 2005; Vermeer, 2000).

5.3.4.3. Accuracy measures

As for the writing task, the **error-free T-units per T-unit (EFTU/T)** ratio was used for oral production, as well as the **errors per T-unit (Err/T)**. These measures have been used in numerous studies exploring accuracy in speaking (Foster & Skehan, 1996; Gilabert, 2005; Mehnert, 1998). Additional reasons to adopt these two measures can be found in section 5.3.3.3.

Summarizing, the following measures have been used in this dissertation in order to analyze both the progress in oral production made by all the students from the beginning until the end of their respective EFL courses, and to determine whether time distribution of EFL instruction affects such progress:

- **Fluency**: syllables per minute (calculated in terms of the Rate B)
- **Complexity**:
    - ***Syntactic***: clauses per T-unit (C/T) (number of clauses divided by the total number of T-units)

- ***Lexical***: Guiraud's Index (number of word types divided by the square root of the number of word tokens)

- **Accuracy**:
  - Percentage of error-free T-units (EFTU/T) (number of error-free T-units divided by total number of T-units x 100)
  - Errors per T-unit (Err/T) (number of errors divided by the total number of T-units)

5.3.4.4. Coding and statistical analyses

The oral data was collected using a tape recorder, which was placed next to the student as she/he was performing the narrative task. Then, transcriptions were made (using a transcriber) in the CHA mode (MacWhinney, 2000). After the samples were transcribed, the coding took place. This researcher was in charge of the collection of the oral data for approximately 75% of all the students (the remaining 25% was completed by four research assistants), as well as the transcription and coding. A research assistant scored a random 10% of the sample, with whom the interrater agreement reached 94%. Intrarater reliability reached 95%.

When all the coding was done, the task was timed using a stopwatch in order to determine the duration of each narrative for the fluency measure (syll/min). Finally, the syllables were counted and added, using a calculator. The CLAN program was used in order to analyze the frequencies of each of the units

considered for the oral samples (total number of T-units, clauses, error-free T-units and errors). Once those frequencies were obtained, the results were copied to an SPSS matrix, as well as the count of syllables and seconds, and the formulas were calculated for the specific measures (syll/min, C/T, Guiraud's Index, EFTU/T, Err/T). Afterwards, all the relevant statistical analyses were executed using SPSS.

CHAPTER 6: RESULTS

This chapter will present the results obtained from the statistical analyses for each of the tasks that the students performed. First, the results of the level 3 learners will be introduced for the written test (section 6.1.1), the writing task (section 6.1.2) and then the speaking task (section 6.1.3). Next, the results obtained by the level 5 students will be presented following the same order: written test (section 6.2.1), writing task (section 6.2.2) and oral production task (section 6.2.3). Section 6.4 will present comparisons of written and oral performance by intermediate and advanced learners. Finally, section 6.5 will analyze the advanced learners' written and oral production as compared to native English speakers.

6.1. Level 3

6.1.1. Students' progress in listening, grammar, reading and vocabulary

There were a total of 131 students in level 3 who completed the listening, sentence conversion, cloze and reading activities. Out of the 131 learners, 49 were in the extensive program, 44 in the semi-intensive, and 38 in the intensive. Table 6.1 presents a summary of the descriptive statistics for each test and each group, including the mean, standard deviation (SD), skewness (Sk) and kurtosis (Kur). In order to examine whether there were significant differences between the three programs in relation to the pretest and the posttest, several two-way mixed

ANOVAs were performed, since this is the statistical analysis which better captures such differences, due to the fact that it simultaneously considers the two variables (type of program and time). In this section, the findings from such tests will be presented, including a table (Table 6.2) which summarizes all the results. Nevertheless, since some measures failed the test of normality and some of them the Levene's Test of Equality of Error Variances, non-parametric tests were performed for those measures (see Appendix H for a detailed account of the results of all these tests). It was observed, though, that the results obtained through parametric and non-parametric tests were virtually the same.

Table 6.1: Descriptive Statistics Written Test Level 3

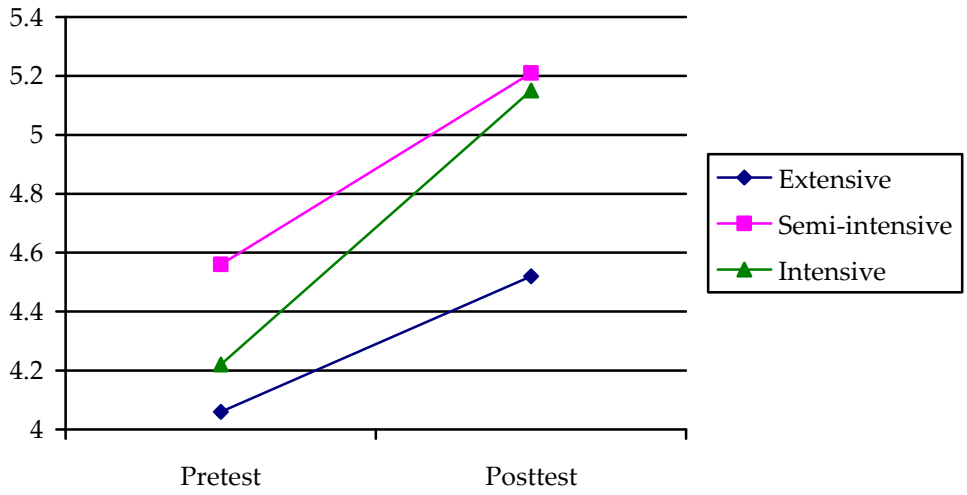
		EXTENSIVE				SEMI-INTENSIVE				INTENSIVE			
		Mean	**SD**	**Sk**	**Kur**	**Mean**	**SD**	**Sk**	**Kur**	**Mean**	**SD**	**Sk**	**Kur**
PRETEST	**List /6**	4.06	1.42	.-505	.020	4.55	1.28	-1.16	2.27	4.22	1.30	-.406	-.253
	SC /5	.694	.865	1.23	.657	.932	1.11	1.21	.469	.829	.983	2.21	7.70
	Cloze /5	2.42	1.06	.077	-.548	2.57	.956	-.434	-.043	2.71	.984	-.449	-.600
	ReadA /5	3.17	.733	-.058	-.263	3.36	.718	-.284	.337	3.33	.700	-.156	-.143
	ReadB /5	1.45	1.12	.412	-.917	1.82	1.17	.095	-.901	1.41	1.16	.440	-.926
	ReadTotal /10	4.62	1.42	.291	-.450	5.18	1.62	.029	-.079	4.74	1.47	.081	-.704
	TOTAL /26	11.80	3.23	.001	-.842	13.24	3.65	.072	-.676	12.50	3.48	-.140	.508
POSTTEST	**List /6**	4.53	1.39	-.547	-.658	5.22	.973	-1.36	1.66	5.15	.968	-1.54	2.73
	SC /5	1.06	1.10	.799	-.369	1.52	1.34	.395	-1.15	1.72	1.34	.762	-.297
	Cloze /5	2.94	1.13	-.417	-.565	3.28	.936	-.802	.170	3.35	.937	-.596	-.521
	ReadA /5	3.50	.721	-.031	-.627	3.86	.681	-.617	.293	3.63	.665	-.295	-.263
	ReadB /5	1.98	1.05	.829	-.010	2.29	1.69	.295	-1.29	2.60	1.08	-.083	-.116
	ReadTotal /10	5.50	1.43	.469	-.629	6.15	2.07	.094	-1.16	6.24	1.37	-.212	-.135
	TOTAL /26	14.03	3.66	.017	-.869	16.18	4.09	-.110	-1.17	16.46	3.44	-.093	-.412

The two-way mixed ANOVA with the **scores on the listening test** as the dependent variable revealed a significant main effect for time ($F(1,128) = 40.07$, $p = .001$, partial $\eta^2 = .238$). There was also a significant main effect for program ($F(2,128) = 3.61$, $p = .030$, partial $\eta^2 = .053$). However, no significant interaction effect was found for program by time ($F(2,128) = 1.52$, $p = .223$, partial $\eta^2 = .023$).

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there were no differences between the three groups in the pretest ($F(2,128) = 1.62$, $p = .202$, partial $\eta^2 = .025$), yet the three types of programs under consideration behaved differently in the posttest ($F(2,128) = 5.14$, $p = .007$, partial $\eta^2 = .074$). The students in the semi-intensive and the intensive program had significantly higher scores than their peers in the extensive program ($p = .013$ and $p = .038$ respectively), whereas no significant differences were found between the two types of intensive courses.

Pairwise comparisons were run, again applying the Bonferroni adjustment. Contrasts revealed that the students in the extensive, semi-intensive and intensive programs made significant progress from pretest to posttest ($F(1,128) = 6.99$, $p = .009$, partial $\eta^2 = .052$; $F(1,128) = 12.64$, $p = .001$, partial $\eta^2 = .090$ and $F(1,128) = 21.63$, $p = .000$, partial $\eta^2 = .145$ respectively). See Figure 6.1.

Figure 6.1: Listening Test Level 3

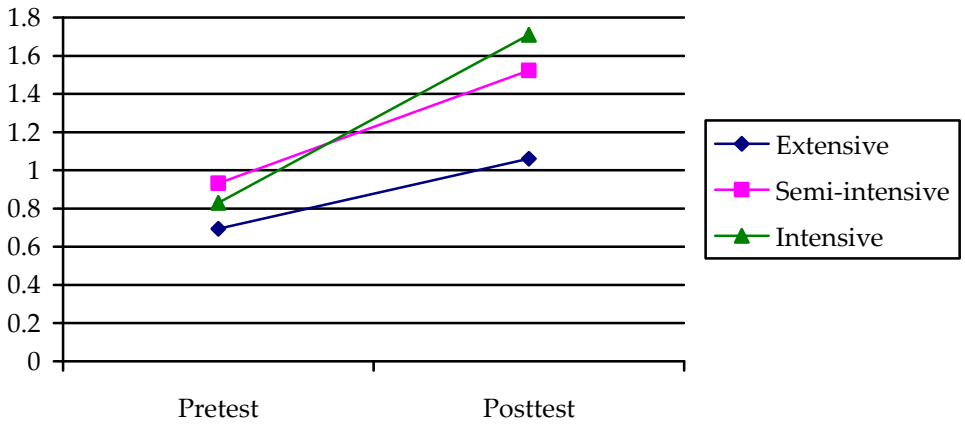


The two-way mixed ANOVA revealed a significant main effect for time in the **sentence conversion exercise** ($F(1,128) = 51.72$, $p = .000$, partial $\eta^2 = .288$). There was no significant main effect for program ($F(2,128) = 2.06$, $p = .132$, partial $\eta^2 = .031$); nevertheless, a significant interaction effect was found for program by time ($F(2,128) = 3.06$, $p = .050$, partial $\eta^2 = .046$), indicating that the programs progressed in time in a significantly different way (see Figure 6.2 for a visual representation).

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there were no differences between the three groups in the pretest ($F(2,128) = .682$, $p = .507$, partial $\eta^2 = .011$), yet the results in the posttest were significantly different ($F(2,128) = 3.21$, $p = .044$, partial $\eta^2 = .048$). At that time, the students in the intensive course obtained significantly higher scores than their peers in the extensive program ($p = .05$), with no other statistically significant differences being found.

Pairwise comparisons applying the Bonferroni adjustment demonstrated that all the groups made significant progress from pretest to posttest (extensive: $F(1,128) = 6.97$, $p = .009$, partial $\eta^2 = .052$; semi-intensive: $F(1,128) = 16.19$, $p = .000$, partial $\eta^2 = .112$ and intensive: $F(1,128) = 31.58$, $p = .000$, partial $\eta^2 = .198$ (see Figure 6.2).

Figure 6.2: Sentence Conversion Level 3



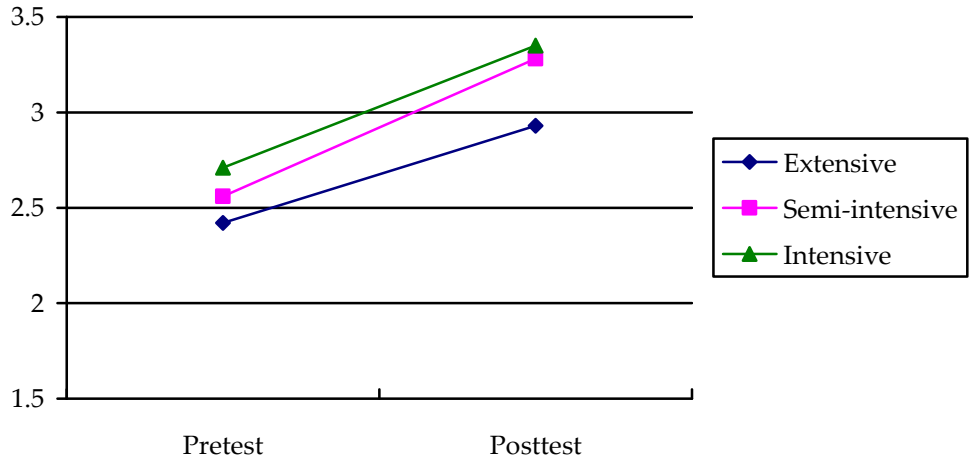
The two-way ANOVA performed with the scores on **the cloze test** as the dependent variable revealed a significant main effect for time ($F(1,128) = 59.53$, $p = .000$, partial $\eta^2 = .317$), but there was no significant main effect for program ($F(1,128) = 1.79$, $p = .170$, partial $\eta^2 = .027$). Neither was the interaction effect for program by time significant ($F(2,128) = .567$, $p = .569$, partial $\eta^2 = .009$).

As with the other measures, an investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there were no differences between the three groups in the pretest ($F(2,128) = .879$, $p =$

.418, partial η^2 = .014). However, as opposed to what was reported for the previous exercises, there were no significant differences between the three types of programs in the posttest (F(2,128) = 2.19, p = .116, partial η^2 = .033).

Pairwise comparisons applying the Bonferroni adjustment showed that the students enrolled in extensive, semi-intensive and intensive programs made significant progress from the pretest to the posttest in the cloze activity (F(1,128) = 15.29, p = .000, partial η^2 = .107; F(1,128) = 26.49, p = .000, partial η^2 = .171; and F(1,128) = 18.56, p = .000, partial η^2 = .127). See Figure 6.3.

Figure 6.3: Results of the Cloze Exercise for the Three Programs



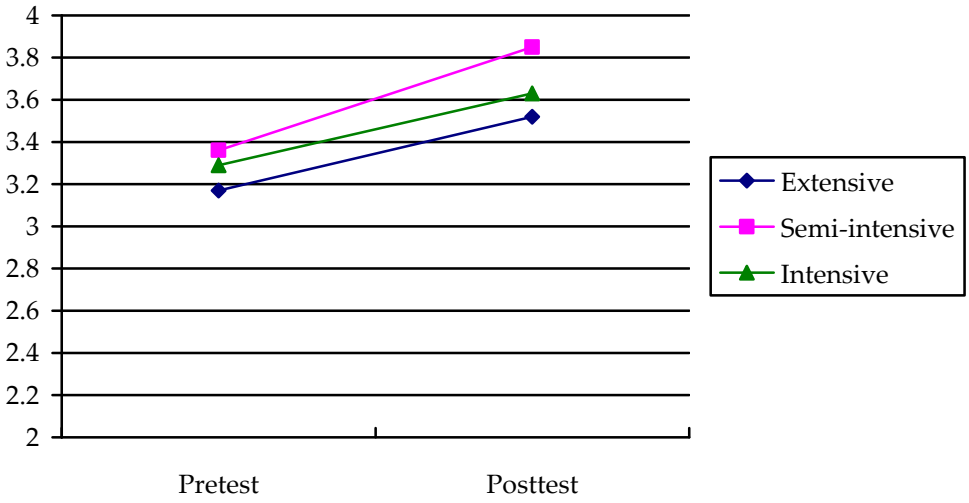
The two-way mixed ANOVA with the **scores on the reading exercise A** (which measured overall comprehension) as the dependent variable revealed a significant main effect for time (F(1,128) = 24.88, p = .000, partial η^2 = .163). There was a marginally significant effect for program (F(2,128) = 2.62, p = .077, partial η^2 = .039). On the other hand, no significant interaction effect was found for program by

time ($F(2,128) = .564$, $p = .570$, partial $\eta^2 = .009$).

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there were no differences between the three groups in the pretest ($F(2,128) = .924$, $p = .400$, partial $\eta^2 = .014$); however, the three types of programs under consideration behaved differently in the posttest, and such difference was close to statistical significance ($F(2,128) = 2.83$, $p = .063$, partial $\eta^2 = .042$), with the students in the semi-intensive course evidencing a marginal statistically significant advantage with respect to the students in the extensive group ($p = .061$).

Results from the Pairwise comparisons applying the Bonferroni adjustment revealed that the students in the extensive, semi-intensive and intensive programs made significant progress from pretest to posttest in the reading A exercise ($F(1,128) = 7.79$, $p = .006$, partial $\eta^2 = .057$; $F(1,128) = 14.20$, $p = .000$, partial $\eta^2 = .100$ and $F(1,128) = 4.60$, $p = .034$, partial $\eta^2 = .035$ respectively). See Figure 6.4.

Figure 6.4: Reading A Level 3



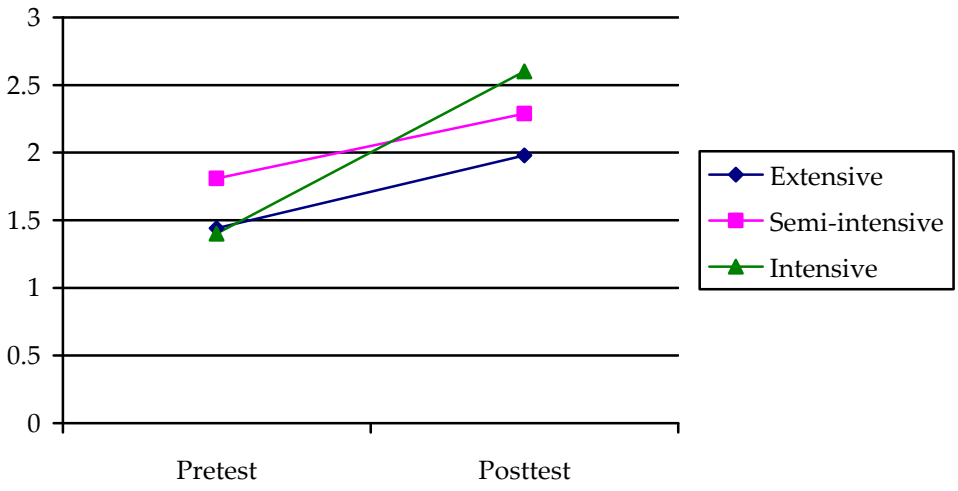
Regarding the **scores on the reading exercise B** (which more clearly measured vocabulary), there was a significant main effect for time ($F(1,128) = 44.60$, $p = .000$, partial $\eta^2 = .258$). Nevertheless, there was no significant main effect for program ($F(2,128) = 1.42$, $p = .247$, partial $\eta^2 = .022$). The interaction effect found for program by time was significant ($F(2,128) = 4.13$, $p = .018$, partial $\eta^2 = .061$), suggesting that the students registered in the three types of programs under analysis progressed in a significantly different manner (see Figure 6.5).

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there were no differences between the three groups in the pretest ($F(2,128) = 1.68$, $p = .191$, partial $\eta^2 = .026$); differences were marginally significant in the posttest ($F(2,128) = 2.46$, $p = .089$, partial $\eta^2 = .037$).

Pairwise comparisons were run, again applying the Bonferroni adjustment.

Contrasts revealed that the students in the extensive, semi-intensive and intensive programs made significant progress from pretest to posttest ($F(1,128) = 8.79$, $p = .004$, partial $\eta^2 = .064$; $F(1,128) = 6.38$, $p = .013$, partial $\eta^2 = .048$ and $F(1,128) = 34.70$, $p = .000$, partial $\eta^2 = .213$ respectively). See Figure 6.5.

Figure 6.5: Reading B Level 3



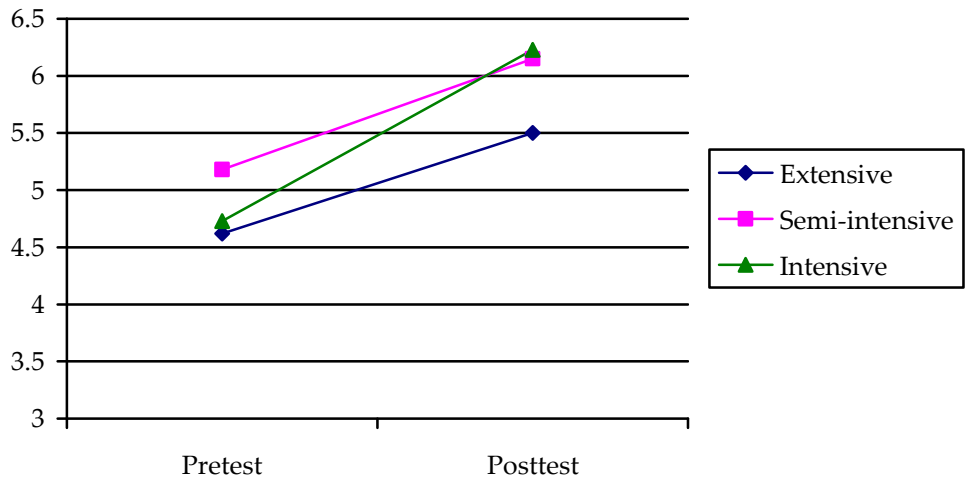
With respect to the **scores on the reading exercise as a whole**, the ANOVA performed showed that there was a significant main effect for time ($F(1,128) = 61.89$, $p = .000$, partial $\eta^2 = .326$) and there was a marginally significant main effect for program ($F(1,128) = 2.44$ $p = .091$, partial $\eta^2 = .037$). No interaction effect between time and program was found ($F(2,128) = 1.77$, $p = .175$, partial $\eta^2 = .027$).

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, demonstrated that there were no differences between the three groups in the pretest ($F(2,128) = .1.73$, $p = .181$, partial $\eta^2 = .026$),

yet in the posttest differences were marginally significant ($F(2,128) = 2.70$, $p = .071$, partial $\eta^2 = .040$).

Contrasts of the Pairwise comparisons using the Bonferroni adjustment revealed that the students in the three programs (extensive, semi-intensive and intensive) made significant progress from the pretest to the posttest ($F(1,128) = 14.46$, $p = .000$, partial $\eta^2 = .101$; $F(1,128) = 15.92$, $p = .000$, partial $\eta^2 = .111$; and $F(1,128) = 32.76$, $p = .000$, partial $\eta^2 = .204$ respectively). See Figure 6.6.

Figure 6.6: Reading Total Level 3



A two-way mixed ANOVA with the **total scores on the written test** as the dependent variable revealed a significant main effect for time ($F(1,128) = 158.0$, $p = .000$, partial $\eta^2 = .552$). There was also a significant main effect for program ($F(2,128) = 3.99$, $p = .021$, partial $\eta^2 = .059$). Additionally, a significant interaction effect was found for program by time ($F(2,128) = 4.26$, $p = .016$, partial $\eta^2 = .062$).

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there were no differences between the three groups in the pretest ($F(2,128) = 2.02$, $p = .137$, partial $\eta^2 = .031$). In contrast, the results of the posttest were significantly different between the three types of programs considered ($F(2,128) = 5.73$, $p = .004$, partial $\eta^2 = .082$). The scores of the students in the extensive group were significantly lower than those of their peers in the semi-intensive ($p = .020$) and the intensive ($p = .010$) programs, while no significant differences existed between the last two programs.

Pairwise comparisons were run, applying the Bonferroni adjustment, which demonstrated, again, that students in the three program types made significant progress from pretest to posttest ($F(1,128) = 31.97$, $p = .000$, partial $\eta^2 = .200$ (extensive); $F(1,128) = 50.06$, $p = .000$, partial $\eta^2 = .281$ (semi-intensive); and $F(1,128) = 78.60$, $p = .000$, partial $\eta^2 = .380$, for the intensive program). See Figure 6.7.

Figure 6.7: Total Written Test Level 3

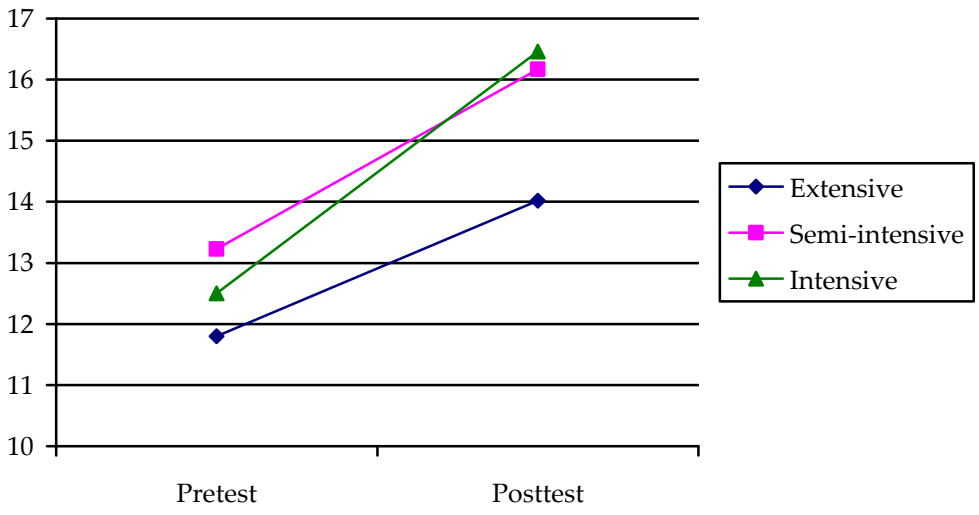


Table 6.2: Results Two-Way Mixed ANOVA Written Test Level 3

	Time	Program	Program x Time	Pretest	Posttest	Pairwise comparisons		
						Extensive	**Semi-intensive**	**Intensive**
Listening	$F(1,128) = 40.07$ $\mathbf{p = .000}$ partial $\eta^2 = .238$	$F(2,128) = 3.61$ $\mathbf{p = .030}$ partial $\eta^2 = .05$	$F(2,128) = 1.52$ $p = .223$ partial $\eta^2 = .023$	$F(2,128) = 1.62$ $p = .202$ partial $\eta^2 = .025$	$F(2,128) = 5.14$ $\mathbf{p = .007}$ partial $\eta^2 = .074$	$F(1,128) = 6.99$ $\mathbf{p = .009}$ partial $\eta^2 = .052$	$F(1,128) = 12.64$ $\mathbf{p = .001}$ partial $\eta^2 = .090$	$F(1,128) = 21.63$ $\mathbf{p = .000}$ partial $\eta^2 = .145$
Sentence conversion	$F(1,128) = 51.72$ $\mathbf{p = .000}$ partial $\eta^2 = .288$	$F(2,128) = 2.06$ $p = .132$ partial $\eta^2 = .031$	$F(2,128) = 3.06$ $\mathbf{p = .050}$ partial $\eta^2 = .046$	$F(2,128) = .682$ $p = .507$ partial $\eta^2 = .011$	$F(2,128) = 3.21$ $\mathbf{p = .044}$ partial $\eta^2 = .048$	$F(1,128) = 6.97$ $\mathbf{p = .009}$ partial $\eta^2 = .052$	$F(1,128) = 16.19$ $\mathbf{p = .000}$ partial $\eta^2 = .112$	$F(1,128) = 31.58$ $\mathbf{p = .000}$ partial $\eta^2 = .198$
Cloze	$F(1,128) = 59.53$ $\mathbf{p = .000}$ partial $\eta^2 = .317$	$F(2,128) = 1.79$ $p = .170$ partial $\eta^2 = .027$	$F(2,128) = .567$ $p = .569$ partial $\eta^2 = .009$	$F(2,128) = .879$ $p = .418$ partial $\eta^2 = .014$	$F(2,128) = 2.19$ $p = .116$ partial $\eta^2 = .033$	$F(1,128) = 15.29$ $\mathbf{p = .000}$ partial $\eta^2 = .107$	$F(1,128) = 26.49$ $\mathbf{p = .000}$ partial $\eta^2 = .171$	$F(1,128) = 18.56$ $\mathbf{p = .000}$ partial $\eta^2 = .127$
Reading A (comprehension)	$F(1,128) = 24.88$ $\mathbf{p = .000}$ partial $\eta^2 = .163$	$F(2,128) = 2.62$ $p = .077$ partial $\eta^2 = .039$	$F(2,128) = .564$ $p = .570$ partial $\eta^2 = .009$	$F(2,128) = .924$ $p = .400$ partial $\eta^2 = .014$	$F(2,128) = 2.83$ $p = .063$ partial $\eta^2 = .04$	$F(1,128) = 7.79$ $\mathbf{p = .006}$ partial $\eta^2 = .057$	$F(1,128) = 14.20$ $\mathbf{p = .000}$ partial $\eta^2 = .100$	$F(1,128) = 4.60$ $\mathbf{p = .034}$ partial $\eta^2 = .035$
Reading B (vocabulary)	$F(1,128) = 44.60$ $\mathbf{p = .000}$ partial $\eta^2 = .258$	$F(2,128) = 1.42$ $p = .247$ partial $\eta^2 = .022$	$F(2,128) = 4.13$ $\mathbf{p = .018}$ partial $\eta^2 = .061$	$F(2,128) = 1.68$ $p = .191$ partial $\eta^2 = .026$	$F(2,128) = 2.46$ $p = .089$ partial $\eta^2 = .037$	$F(1,128) = 8.79$ $\mathbf{p = .004}$ partial $\eta^2 = .064$	$F(1,128) = 6.38$ $\mathbf{p = .013}$ partial $\eta^2 = .048$	$F(1,128) = 34.70$ $\mathbf{p = .000}$ partial $\eta^2 = .213$
Reading Total	$F(1,128) = 61.89$ $\mathbf{p = .000}$ partial $\eta^2 = .326$	$F(2,128) = 2.44$ $p = .091$ partial $\eta^2 = .037$	$F(2,128) = 1.77$ $p = .175$ partial $\eta^2 = .027$	$F(2,128) = 1.73$ $p = .181$ partial $\eta^2 = .026$	$F(2,128) = 2.70$ $p = .071$ Partial $\eta^2 = .040$	$F(1,128) = 14.46$ $\mathbf{p = .000}$ partial $\eta^2 = .101$	$F(1,128) = 15.92$ $\mathbf{p = .000}$ partial $\eta^2 = .111$	$F(1,128) = 32.76$ $\mathbf{p = .000}$ Partial $\eta^2 = .204$
TOTAL	$F(1,128) = 158.0$ $\mathbf{p = .000}$ partial $\eta^2 = .552$	$F(2,128) = 3.99$ $\mathbf{p = .021}$ partial $\eta^2 = .059$	$F(2,128) = 4.26$ $\mathbf{p = .016}$ partial $\eta^2 = .062$	$F(2,128) = 2.02$ $p = .137$ partial $\eta^2 = .031$	$F(2,128) = 5.73$ $\mathbf{p = .004}$ partial $\eta^2 = .082$	$F(1,128) = 31.97$ $\mathbf{p = .000}$ partial $\eta^2 = .200$	$F(1,128) = 50.06$ $\mathbf{p = .000}$ partial $\eta^2 = .281$	$F(1,128) = 78.60$ $\mathbf{p = .000}$ partial $\eta^2 = .380$

In **sum**, as can be observed in Table 6.2, there was always a significant effect for time, which means that the results of the pretest were significantly different to those of the posttest for the three programs. Nevertheless, the types of program under research were not always significantly different (only in the listening for the variable *program*, in the sentence conversion and reading B in *program by time* and in the total scores on the test in *program* and *program by time*). No significant differences across programs were found in the cloze, reading A and reading total. It is in the areas where differences among groups existed (listening, sentence conversion, reading A, and total scores) that the results of the posttest were significantly higher for the students in the semi-intensive and the intensive groups than for those enrolled in the extensive course. Moreover, when analyzing the size of the effect as represented in the pairwise comparisons, it can be observed that, while all the groups significantly progressed from pre to posttest, the partial eta squared values were almost always superior, except in the reading A, in the semi-intensive and intensive groups (especially in the latter for all the tasks except for the cloze), which means that the size of the effect was usually larger for the most intensive program.

Therefore, taking into account all these results, it can be concluded that there are some significant interaction effects that relate improvement to intensity, especially in terms of performance in the listening, sentence conversion and reading B tasks. When significant differences exist, they are always in favor of the students in the semi-intensive and the intensive groups.

6.1.2. Students' progress in writing skills

The students' progress in writing was measured in terms of fluency (words per T-unit, W/T), complexity in syntax (clauses per T-unit, C/T) and in vocabulary (Guiraud's Index of Lexical Richness), and accuracy (percentage of error-free T-units per T-unit, EFTU/T; and errors per T-unit, Err/T). The results of the descriptive statistics revealed that there were a number of outliers; as a consequence, the data was not normally distributed for most of the areas tested for all the groups in the study. Other alternatives were used to check for normality, yet the distribution of the sample always appeared to be abnormal, which is why it was decided to eliminate those outliers. Therefore, there will be a lower *n* in these samples; however, the number was always higher than 30 (40 in the extensive, 31 in the semi-intensive, and 32 in the intensive). Table 6.3 presents the descriptive statistics for the final sample. To see the results of the normality tests and the Levene's Test see Appendix I.

Table 6.3: Descriptive Statistics Writing Level 3

		EXTENSIVE				SEMI-INTENSIVE				INTENSIVE			
		Mean	**SD**	**Sk**	**Kur**	**Mean**	**SD**	**Sk**	**Kur**	**Mean**	**SD**	**Sk**	**Kur**
PRETEST	**Fluency (W/T)**	8.71	1.45	.340	-.382	8.71	1.61	-.082	-1.01	8.27	1.20	-.079	-.749
	Syntactic C. (C/T)	1.55	.281	.726	.723	1.51	.298	.468	-.128	1.50	.188	.694	.383
	Lexical C. (Guiraud's I.)	7.26	.716	.674	-.159	7.51	.838	-.324	-.089	7.09	.720	-.105	-.446
	Accuracy 1 (EFTU/T)	51.54	17.78	.192	-.702	57.71	15.67	-.594	.286	56.59	18.50	-.196	-.803
	Accuracy 2 (Err/T)	.756	.396	.539	-.532	.605	.310	1.43	2.23	.649	.370	.738	-.002
POSTTEST	**Fluency (W/T)**	10.85	2.40	.653	-.261	9.23	1.55	.449	-.006	10.36	1.81	.349	-.431
	Syntactic C. (C/T)	1.94	.446	.914	.180	1.63	.320	.289	-.159	1.89	.352	.339	-.472
	Lexical C. (Guiraud's I.)	6.59	.683	.200	-.163	6.94	.747	-.390	-.435	7.49	.794	-.782	.646
	Accuracy 1 (EFTU/T)	44.20	14.89	.120	.024	56.68	17.80	-.203	-.916	51.23	20.20	.327	-.325
	Accuracy 2 (Err/T)	.957	.357	.584	.796	.627	.343	.493	-.647	.766	.435	.597	-.221

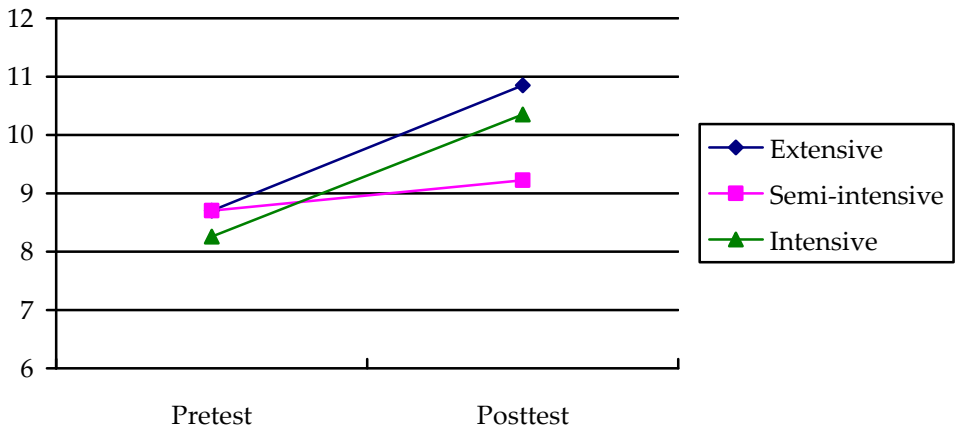
Two-way mixed ANOVAs were performed for all the writing measures in order to investigate the effects of type of program (extensive, semi-intensive and intensive) and time (pretest and posttest).

The results of such test for the **words per T-unit (W/T)** measure revealed that there was a significant main effect for time, program, and program by time ($F(1,100)= 51.19$, $p = .000$, partial $\eta^2 = .339$; $F(2,100)= 3.37$, $p= .038$, partial $\eta^2= .063$; $F(2,100) = 5.56$, $p= .005$, partial $\eta^2= .100$ respectively).

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there were no differences between the three groups in the pretest ($F(1,100) = 1.05$, $p = .335$, partial $\eta^2 = .021$), yet the results of the posttest were significantly different ($F(2,100) = 5.95$, $p = .004$, partial $\eta^2 = .106$), with the students in the extensive program having significantly higher scores than those in the semi-intensive program ($p = .003$).

Pairwise comparisons were run, again applying the Bonferroni adjustment. Contrasts revealed that the students in both, the extensive and the intensive program, used significantly more words per T-unit in the posttest than in the pretest ($F(1,100) = 36.87$, $p = .000$, partial $\eta^2 = .269$, and $F(1,100) = 28.00$, $p = .000$, partial $\eta^2 = .219$, respectively). The semi-intensive group also used more words per T-unit in the posttest, yet no significant difference was obtained in the statistical analysis between the two times ($F(1,100) = 1.68$, $p = .197$, partial $\eta^2 = .017$). See Figure 6.8.

Figure 6.8: Fluency (W/T) Writing Level 3



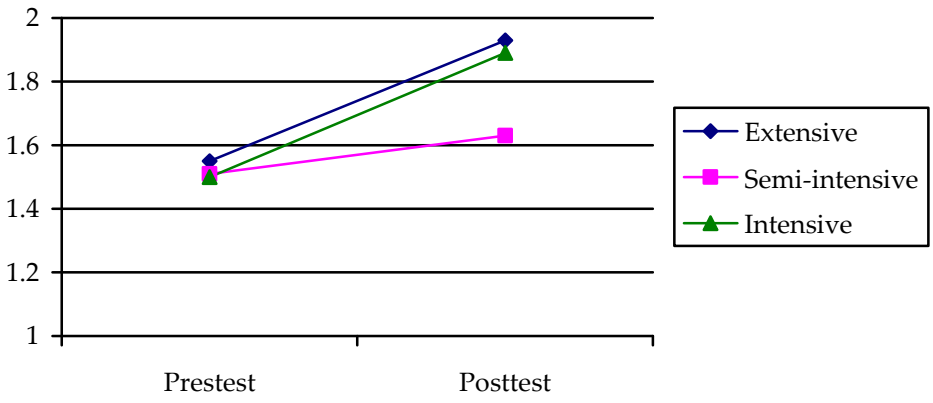
The results of the syntactic complexity measure **clauses per T-unit (C/T)** were quite similar. The analyses revealed a significant main effect of time ($F(1,100) = 44.67$, $p = .000$, partial $\eta^2 = .309$), program ($F(2,100) = 4.76$, $p = .011$, partial $\eta^2 = .087$), and program by time ($F(2,100) = 3.79$, $p = .026$, partial $\eta^2 = .071$).

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there were no differences between the three program types in the pretest ($F(2,100) = .379$, $p = .686$, partial $\eta^2 = .008$); nevertheless, the results of the posttest were significantly different among the three program types ($F(2,100) = 6.13$, $p = .003$, partial $\eta^2 = .109$). The semi-intensive group produced significantly fewer clauses per T-unit than the extensive ($p = .003$) and intensive ($p = .027$) groups.

Pairwise comparisons applying the Bonferroni adjustment revealed again that only the extensive and the intensive groups had used significantly more clauses per T-unit in the pretest than in the posttest ($F(1,100) = 29.54$, $p = .000$, partial $\eta^2 = .228$, and $F(1,100) = 23.71$, $p = .000$, partial $\eta^2 = .192$). The semi-intensive group

also progressed in terms of syntactic complexity, yet such progress was not statistically significant ($F(1,100) = 2.23$, $p = .138$, partial $\eta^2 = .022$). See Figure 6.9.

Figure 6.9: Syntactic Complexity (C/T) Writing Level 3



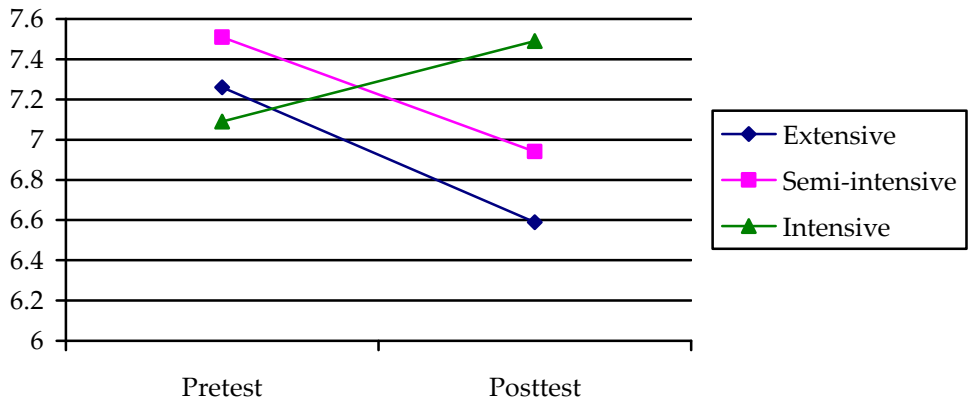
The two-way mixed ANOVA performed with the scores on **the Guiraud's Index of Lexical Richness** as the dependent variable revealed a significant main effect for time ($F(1,100) = 10.02$, $p = .002$, partial $\eta^2 = .091$), program ($F(2,100) = 3.79$, $p = .026$, partial $\eta^2 = .070$), and program by time ($F(2,100) = 14.63$, $p = .000$, partial $\eta^2 = .226$).

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there were no differences between the three groups in the pretest ($F(2,100) = 2.42$, $p = .094$, partial $\eta^2 = .046$), yet the results of the posttest were significantly different ($F(2,100) = 13.12$, $p = .000$, partial $\eta^2 = .208$). The students in the intensive program had a significantly higher index than those in the extensive ($p = .000$) and the semi-intensive ($p = .012$)

programs.

Pairwise comparisons applying the Bonferroni adjustment revealed that the extensive group had a significantly lower score in the posttest than in the pretest ($F(1,100) = 22.56$, $p = .000$, partial $\eta^2 = .184$), the same as the semi-intensive group ($F(1,100) = 12.58$, $p = .001$, partial $\eta^2 = .112$). The students in the intensive group, on the other hand, significantly improved their performance from the pretest to the posttest ($F(1,100) = 6.37$, $p = .013$, partial $\eta^2 = .060$). See Figure 6.10.

Figure 6.10: Guiraud´s Index Writing Level 3



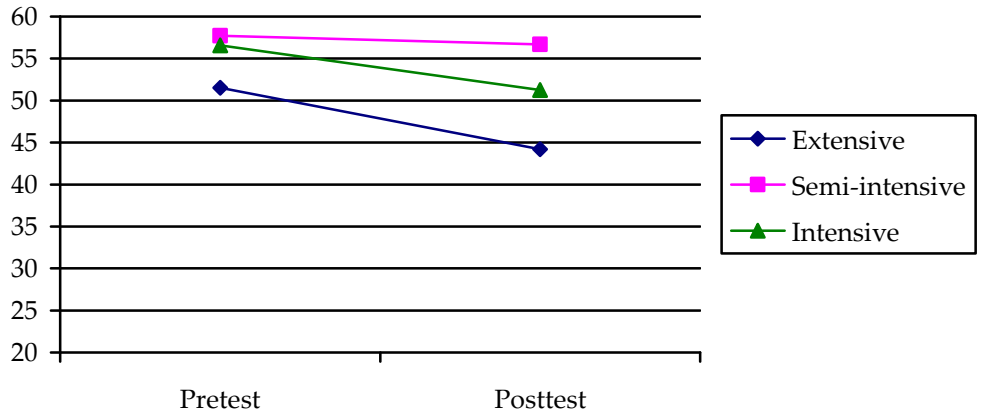
When analyzing the percentage of **error-free T-units per T-unit (EFTU/T)**, the ANOVA performed showed that there was a significant main effect for time ($F(1,100) = 4.48$, $p = .031$, partial $\eta^2 = .046$) and for program ($F(2,100) = 4.13$, $p = .019$, partial $\eta^2 = .076$). On the other hand, there was no significant interaction effect between time and program ($F(2,100) = .793$, $p = .456$, partial $\eta^2 = .016$).

An investigation of the simple main effects, applying the Bonferroni

adjustment for multiple comparison, demonstrated that there were no differences between the three groups in the pretest ($F(2,100) = 1.30$, $p = .277$, partial $\eta^2 = .025$), yet there were significant differences in the posttest ($F(2,100) = 4.50$, $p = .013$, partial $\eta^2 = .083$). In this case, the students in the extensive program had significantly fewer EFTU/T than those in the semi-intensive ($p = .011$).

Pairwise comparisons using the Bonferroni adjustment revealed that the students in the semi-intensive and the intensive programs used fewer error-free T-units per T-unit in the posttest; however, the difference with the pretest was not statistically significant ($F(1,100) = .074$, $p = .786$, partial $\eta^2 = .001$, and $F(1,100) = 2.06$, $p = .154$, partial $\eta^2 = .020$, respectively). In contrast, the students in the extensive program produced significantly fewer error-free T-units per T-unit in the posttest ($F(1,100) = 4.82$, $p = .030$, partial $\eta^2 = .046$). See Figure 6.11.

Figure 6.11: Accuracy (EFTU/T) Writing Level 3



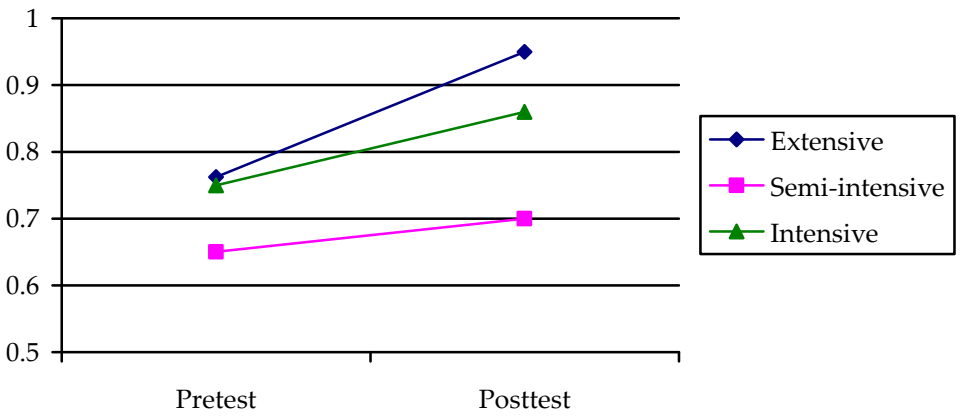
Finally, the outcome of the two-way mixed ANOVA with the **errors per T-**

unit (Err/T) measure as the dependent variable was highly similar to the one reported for the other accuracy measure used in this study (EFTU/T). The results revealed a significant main effect for time ($F(1,100) = 6.97$, $p = .010$, partial $\eta^2 = .065$) and for program ($F(2,100) = 5.79$, $p = .004$, partial $\eta^2 = .104$). No significant interaction effect was found for program by time ($F(2,100) = 1.47$, $p = .235$, partial $\eta^2 = .029$).

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there were no differences between the three groups in the pretest ($F(2,100) = 1.66$, $p = .195$, partial $\eta^2 = .032$); nevertheless, the results in the posttest were significantly different across groups ($F(2,100) = 6.76$, $p = .002$, partial $\eta^2 = .119$), with the students in the extensive program making significantly more errors in the posttest than their peers in the semi-intensive course ($p = .001$).

Pairwise comparisons were run, again applying the Bonferroni adjustment. Contrasts revealed that whereas the extensive group made significantly more errors per T-unit in the posttest ($F(1,100) = 8.55$, $p = .004$, partial $\eta^2 = .079$), such significance was not found in the case of the semi-intensive ($F(1,100) = .085$, $p = .771$, partial $\eta^2 = .001$) or intensive program ($F(1,100) = 2.35$, $p = .129$, partial $\eta^2 = .023$), even if both groups also made more errors in the posttest. See Figure 6.12.

Figure 6.12: Accuracy (Err/T) Writing Level 3



In general, with respect to the variable *program*, significant differences were obtained, as shown in the columns *program* and *program by time* in Table 6.4. It can be claimed that the students in the extensive and the intensive program demonstrated a more significant progress in terms of fluency and syntactic complexity than the learners in the semi-intensive group. Moreover, the students in the intensive program show significantly more complex vocabulary than their peers in the other two groups. Concerning accuracy, the extensive program differed significantly from the other two, because the students in such program, despite progressing in fluency and complexity, had significantly lower scores in accuracy in the posttest.

Regarding *time* there was a change from pre to posttest in all the measures. It can be said that all the students progressed in terms of fluency and syntactic complexity (significantly in the case of the extensive and intensive programs, as seen in the 'pairwise comparisons' Table 6.4), maybe at the expense of accuracy,

which was worse in the posttest than in the pretest for all the students and significantly worse for those in the extensive program (see Table 6.4, 'pairwise comparisons'). Something similar can be said about lexical complexity, which was worse in the posttest than in the posttest for all the groups except for the intensive, for which lexical complexity improved significantly from pre to posttest.

In sum, the program which demonstrated more improvement in writing, considering the four areas under examination, is the intensive, since the students in that program significantly progressed in three of the four areas analyzed (fluency, syntactic complexity, and lexical richness) without such progress having significantly negative effects on accuracy. The semi-intensive program, on the other hand, did not show any significant positive or negative change from pre to posttest. Finally, those students in the extensive program significantly progressed in some measures of writing proficiency (fluency and syntactic complexity) while significantly obtaining lower results in others (accuracy and lexical complexity).

Table 6.4: Results Two-Way Mixed ANOVA Writing Level 3

	Time	Program	Program x Time	Pretest	Posttest	Pairwise comparisons		
						Extensive	Semi-intensive	Intensive
Fluency (W/T)	$F(1,100) = 51.19$ **$p = .000$** partial $\eta^2 = .339$	$F(2,100) = 3.37$ **$p = .038$** partial $\eta^2 = .063$	$F(2,100) = 5.56$ **$p = .005$** partial $\eta^2 = .100$	$F(2,100) = 1.05$ $p = .335$ partial $\eta^2 = .021$	$F(2,100) = 5.95$ **$p = .004$** partial $\eta^2 = .106$	$F(1,100) = 36.87$ **$p = .000$** partial $\eta^2 = .269$	$F(1,100) = 1.68$ $p = .197$ partial $\eta^2 = .017$	$F(1,100) = 28.00$ **$p = .000$** partial $\eta^2 = .219$
Syntactic Complexity (C/T)	$F(1,100) = 44.67$ **$p = .000$** partial $\eta^2 = .309$	$F(2,100) = 4.76$ **$p = .011$** partial $\eta^2 = .087$	$F(2,100) = 3.79$ **$p = .026$** partial $\eta^2 = .071$	$F(2,100) = .379$ $p = .686$ partial $\eta^2 = .008$	$F(2,100) = 6.13$ **$p = .003$** partial $\eta^2 = .109$	$F(1,100) = 29.54$ **$p = .000$** partial $\eta^2 = .228$	$F(1,100) = 2.23$ $p = .138$ partial $\eta^2 = .022$	$F(1,100) = 23.71$ **$p = .000$** partial $\eta^2 = .192$
Lexical richness (Guiraud's I.)	$F(1,100) = 10.02$ **$p = .002$** partial $\eta^2 = .091$	$F(2,100) = 3.79$ **$p = .026$** partial $\eta^2 = .070$	$F(2,100) = 14.63$ **$p = .000$** partial $\eta^2 = .226$	$F(2,100) = 2.42$ $p = .094$ partial $\eta^2 = .046$	$F(2,100) = 13.12$ **$p = .000$** partial $\eta^2 = .208$	$F(1,100) = 22.56$ **$p = .000$** partial $\eta^2 = .184$	$F(1,100) = 12.58$ **$p = .001$** partial $\eta^2 = .112$	$F(1,100) = 6.37$ **$p = .013$** partial $\eta^2 = .060$
Accuracy (EFTU/T)	$F(1,100) = 4.48$ **$p = .031$** partial $\eta^2 = .046$	$F(2,100) = 4.13$ **$p = .019$** partial $\eta^2 = .076$	$F(2,100) = .793$ $p = .456$ partial $\eta^2 = .016$	$F(2,100) = 1.30$ $p = .277$ partial $\eta^2 = .025$	$F(2,100) = 4.50$ **$p = .013$** partial $\eta^2 = .083$	$F(1,100) = 4.83$ **$p = .030$** partial $\eta^2 = .046$	$F(1,100) = .074$ $p = .786$ partial $\eta^2 = .001$	$F(1,100) = 2.06$ $p = .154$ partial $\eta^2 = .020$
Accuracy (Err/T)	$F(1,100) = 6.97$ **$p = .010$** partial $\eta^2 = .065$	$F(2,100) = 5.79$ **$p = .004$** partial $\eta^2 = .104$	$F(2,100) = 1.47$ $p = .235$ partial $\eta^2 = .029$	$F(2,100) = 1.66$ $p = .195$ partial $\eta^2 = .032$	$F(2,100) = 6.76$ **$p = .002$** partial $\eta^2 = .119$	$F(1,100) = 8.55$ **$p = .004$** partial $\eta^2 = .079$	$F(1,100) = .085$ $p = .771$ partial $\eta^2 = .001$	$F(1,100) = 2.35$ $p = .129$ partial $\eta^2 = .023$

6.1.3. Students' progress in speaking skills

The students' progress in speaking skills was measured in terms of fluency (syllables per minute), complexity in syntax (clauses per T-unit) lexical richness (Guiraud's Index), and accuracy (percentage of error-free T-units per T-unit, and errors per T-unit). The students who took the oral test were fewer than those who did the written test, since only a random sample of students from each class was chosen to perform the oral task due to practical reasons. As already reported in Table 5.3, the final numbers are 21 students in the extensive and intensive groups and 11 in the semi-intensive. The normality tests are presented in Appendix J.

Owing to the small n in the three groups, non-parametric tests were considered more adequate to analyze the oral data. In order to examine differences among the three program types for each of the measures considered, Kruskal-Wallis tests were performed, with the three programs as independent variables and each of the measures in the pre and posttest as dependent variables. When significant differences among groups appeared, the Mann-Whitney U Test was performed with only two groups as independent variables in order to analyze differences between one program type and each of the others. Wilcoxon Signed Rank tests were employed to examine differences in the students' performance at time 1 (pretest) and time 2 (posttest) for each of the program types individually. Through this test it was possible to determine whether the progress from the pre to the posttest was significant in the case of extensive, semi-intensive and extensive programs. Table 6.5 shows the descriptive statistics for each measure for each of

the groups under consideration.

The Kruskall-Wallis Test revealed that there were some significant differences among the three programs considered in the **pretest**, with respect to the fluency measure syll/min (χ^2 10.21, *df* 2, p = .006), and in the syntactic complexity measure C/T (χ^2 6.07, *df* 2, p = .048). The students' performance in the rest of the measures in the pretest was not significantly different across groups (Guiraud's Index: χ^2 .571, *df* 2, p = .752; EFTU/T: χ^2 3.00, *df* 2, p = .222, Err/T: χ^2 3.40, *df* 2, p = .183).

Table 6.5: Descriptive Statistics Speaking Level 3

		EXTENSIVE				SEMI- INTENSIVE				INTENSIVE			
		Mean	**SD**	**Sk**	**Kur**	**Mean**	**SD**	**Sk**	**Kur**	**Mean**	**SD**	**Sk**	**Kur**
PRETEST	**Fluency (Syll/min)**	62.00	12.99	-.033	-.260	80.94	12.50	-.150	.091	75.11	23.69	.010	-.484
	Syntactic C. (C/T)	1.49	.403	1.41	1.96	1.35	.217	.169	-.329	1.67	.358	.768	-.048
	Lexical C. (Guiraud's I)	4.94	.560	-.126	-.758	5.06	.479	.149	-1.25	4.91	.566	-.038	-1.26
	Accuracy 1 (EFTU/T)	31.05	16.35	-.333	-1.15	40.03	25.39	-.165	-1.03	26.08	17.12	-.179	-.948
	Accuracy 2 (Err/T)	1.09	.450	.662	-.309	.902	.480	.836	-.627	1.27	.600	1.10	.772
POSTTEST	**Fluency (Syll/min)**	73.54	16.39	.781	1.31	84.29	17.00	-.177	-.866	84.34	17.97	.110	-.859
	Syntactic C. (C/T)	1.66	.365	.061	-1.21	1.51	.204	-.746	1.30	1.68	.385	.750	-.688
	Lexical C. (Guiraud's I)	4.96	.582	-.333	.041	5.03	.505	-.202	-.550	5.17	.470	.084	1.12
	Accuracy 1 (EFTU/T)	35.64	20.27	.770	.000	44.46	15.99	-.137	-.839	38.76	21.93	.286	-.809
	Accuracy 2 (Err/T)	1.15	.393	.181	.-550	.969	.394	.777	-.169	.989	.543	1.48	3.30

In order to find out which groups were significantly different from each other in terms of fluency in the pretest, Mann-Whitney *U* Tests were performed with the fluency and syntactic complexity measures in the pretest as dependent variables and the extensive and semi-intensive groups as the independent variables. The differences between the two groups were statistically significant in oral fluency (*U* 32, *Z* -3.31, *p* = .001), with the semi-intensive students having higher scores than their peers in the extensive program (see Figure 6.13). The analysis of the clauses per T-unit shows that the students in the two program types were not significantly different in this respect in the pretest (*U* 101.5 *Z* -.557, *p* = .584).

When comparing the students in the extensive and the intensive programs, the Mann-Whitney *U* Test revealed a significant difference between the two program types for the syllables/minute measure (*U* 140, *Z* -2.02, *p* = .043), with the students in the intensive program being more fluent (Figure 6.13); however, there were no significant differences in clauses per T-unit (*U* 154, *Z* -1.67, *p* = .094).

Finally, the comparison between the semi-intensive and intensive programs in the pretest using the Mann-Whitney *U* Test demonstrated that these two groups were not significantly different in terms of fluency (*U* 95, *Z* -.813, *p* = .434), yet, as can bee seen in Figure 6.14, the scores on the clauses per T-unit were significantly higher in the intensive program than in the semi-intensive (*U* 53.5, *Z* -2.46, *p* = .012).

Figure 6.13: Fluency (Syll/Min) Speaking Level 3

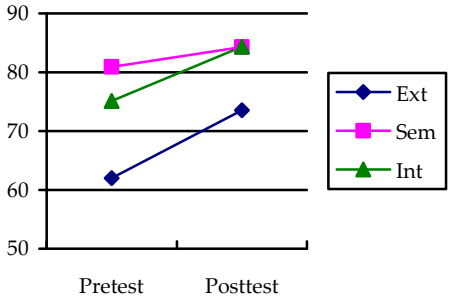


Figure 6.14: Synt. C. (C/T) Sp Level 3

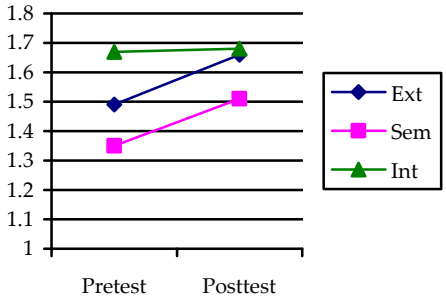


Figure 6.15: Lex. C. (G.I) Sp Level 3

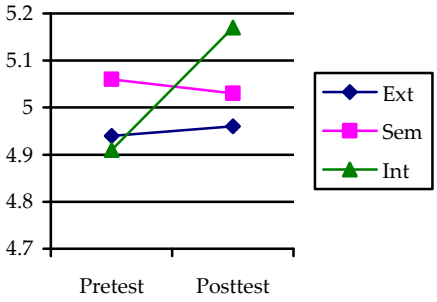


Figure 6.16: Acc. (EFTU/T) Sp Level 3

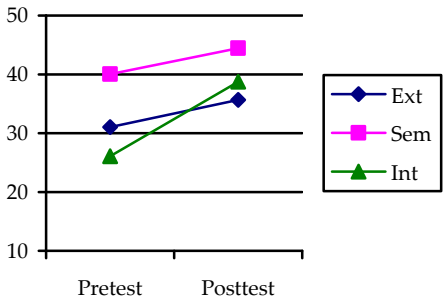
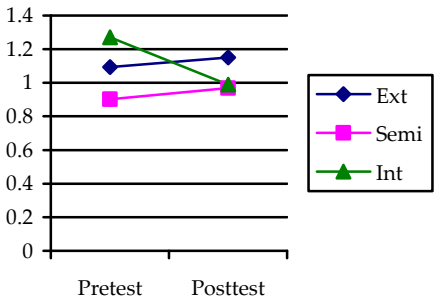


Figure 6.17: Acc. (Err/T) Sp Level 3



Regarding the **posttest**, the Kruskal-Wallis Test revealed that there were no significant differences in any of the measures considered across program types (see Table 6.6).

Table 6.6: Results Kruskal-Wallis Speaking Posttest Level 3

	χ^2	*df*	*P*
Post Syll/min	5.00	2	.082
Post C/T	.965	2	.617
Post Guiraud	1.15	2	.562
Post EFTU/T	1.84	2	.398
Post Err/T	3.32	2	.190

These results indicate that the advantage that the students in the semi-intensive and intensive programs had in the pretest with respect to fluency and syntactic complexity respectively was not maintained at the end of the course. Nevertheless, the students in those programs were always superior to their peers in the extensive program, and, although the differences were not significant in the posttest, the *p* value for the fluency measure is close to statistical significance, as shown in Table 6.6.

In order to examine the *time* factor, Wilcoxon Signed Rank Tests were performed for each of the groups (extensive, semi-intensive and intensive) for each of the oral production measures in the pretest and in the posttest. The results of such tests revealed that, in the posttest, the students in the extensive program significantly improved their fluency (*Z* -3.32, *p* = .001) and syntactic complexity (*Z* -2.05, *p* = .040) with respect to their oral performance. On the other hand, no significant improvement was found for lexical complexity (*Z* -.330, *p* = .741), or accuracy (EFTU/T: *Z* -.852, *p* = .394, Err/T: *Z* -.817, *p* = .414).

The results of the Wilcoxon Signed Rank Test performed for the students in the semi-intensive program with the measures considered to examine oral production showed that the students in such program significantly improved their speaking skills from pre to posttest with respect to syntactic complexity (Z -2.31, p = .021); nevertheless, no significant improvement was registered in fluency (Z -1.33, p = .182), lexical complexity (Z -.178, p = .859), or accuracy (EFTU/T: Z -.889, p = .374; Err/T: Z -.889, p = .374).

According to the same test, the students in the intensive group made significant progress in their speaking skills in all the areas of oral production examined, except for syntactic complexity. The results of the Wilcoxon Signed Rank Test indicate that the students in the intensive group improved their fluency (Z -1.96, p = .050), their lexical complexity (Z -1.99, p = .046), and their accuracy (EFTU/T: Z -2.58, p = .010; Err/T: Z -2.21, p = .027). Nevertheless, no similar improvement was experienced in syntactic complexity (Z -.112, p = .911).

Summarizing, when considering the *program* variable, although there were no differences in the posttest, in the pretest the students in the extensive program were significantly less fluent than the students in the other two groups; moreover, the students in the semi-intensive group demonstrated significantly less syntactic complexity than their peers in the intensive course, as indicated by the Kruskal-Wallis Test.

Nevertheless, when examining the *time* variable, certain differences can be found in the way the students in the three programs under study progressed from the pretest to the posttest. To begin with, the students in the extensive group

significantly improved their oral performance in terms of fluency and syntactic complexity, whereas no significant improvement was observed in any of the other areas examined (in fact, accuracy in the posttest is slightly worse than in the pretest in terms of errors per T-unit). The students in the semi-intensive program only improved in a statistically significant way regarding syntactic complexity. Some improvement in fluency occurred, but it was not significant. Slightly worse performance was observed in terms of lexical richness and one accuracy measure (errors per T-unit). Finally, the students in the intensive program made a significant progress in oral production from pre to posttest in all the areas analyzed (fluency, lexical complexity, and accuracy) except for one (syntactic complexity). See Figure 6.13 through Figure 6.17 for a graphic representation of the students' performance.

6.2. Level 5

6.2.1. Students' progress in listening, grammar and vocabulary

Table 6.7 presents the descriptive statistics for the results of the level 5 listening (List), sencence conversion (SC), cloze, and total scores of the written test, including the means (M), standard deviations (SD), skewness (Sk) and kurtosis (Kur). Appendix K provides a detailed account of the normality tests and the Levene's Test of Equality of Error Variance. According to such tests the samples considered have a normal distribution and the error variance of the dependent

variable is equal across groups.

Table 6.7: Descriptive Statistics Written Test Level 5

		EXTENSIVE				INTENSIVE			
		M	**SD**	**Sk**	**Kur**	**M**	**SD**	**Sk**	**Kur**
PRETEST	**List /9**	3.95	1.38	.177	-.639	4.37	1.60	.518	-.613
	SC /5	2.21	1.24	.325	-.461	2.47	1.27	.363	-.080
	Cloze /5	2.43	1.13	-.294	.000	2.74	1.05	.071	-.586
	TOT /19	8.43	2.67	.066	-.254	9.51	2.67	-.085	-.823
POSTTEST	**List /6**	4.72	1.38	-.239	.354	5.23	1.58	-.180	-.534
	SC /5	2.40	1.35	.089	-.832	2.51	1.25	-.048	-.034
	Cloze /5	3.12	.869	-.028	-.878	3.29	.834	-.472	.038
	TOT /19	10.15	2.63	-.052	-.387	11.03	2.54	-.400	-.497

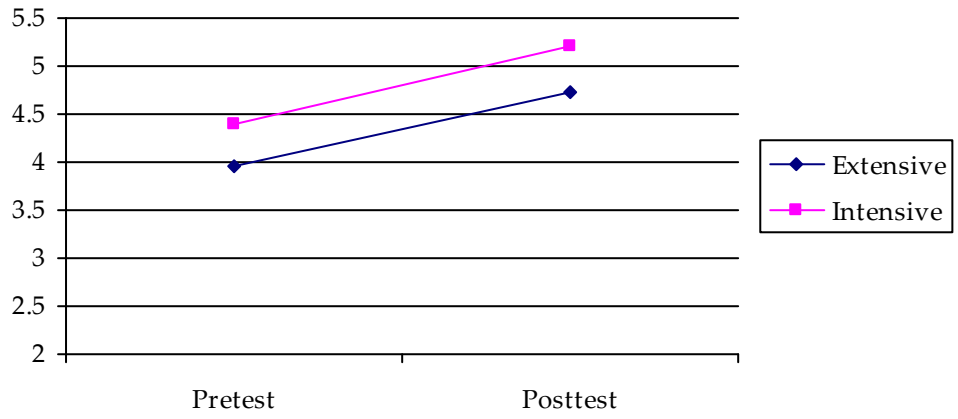
Since it can be assumed that the samples for each of the measures in the pre and posttest for both types of programs under research have a normal distribution and include a number of more than 30 subjects (34 in the extensive group, 31 in the intensive), the parametric test two-way ANOVA can be applied without any of the concerns which appeared for some exercises within the level 3 written test.

The two-way mixed ANOVA with the **scores on the listening test** as the dependent variable revealed a significant main effect for time ($F(1,66) = 26.53$ $p = .000$, partial $\eta^2 = .264$). However, there was no significant main effect for program ($F(1,66) = 2.15$, $p = .147$, partial $\eta^2 = .032$), or for program by time ($F(1,66) = .065$, $p = .800$, partial $\eta^2 = .001$).

No differences between the two groups were found in the pretest nor in the posttest, according to the investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison ($F(1,66) = 1.38$, $p = .244$, partial $\eta^2 = .020$; $F(1,66) = 2.01$, $p = .160$, partial $\eta^2 = .030$, respectively).

Pairwise comparisons were run, applying the Bonferroni adjustment. Contrasts revealed that the students in the extensive and intensive programs made significant progress from pretest to posttest ($F(1,66) = 11.64$, $p = .001$, partial $\eta^2 = .150$; $F(1,66) = 12.04$, $p = .001$, partial $\eta^2 = .154$ respectively). See Figure 6.18.

Figure 6.18: Listening Level 5

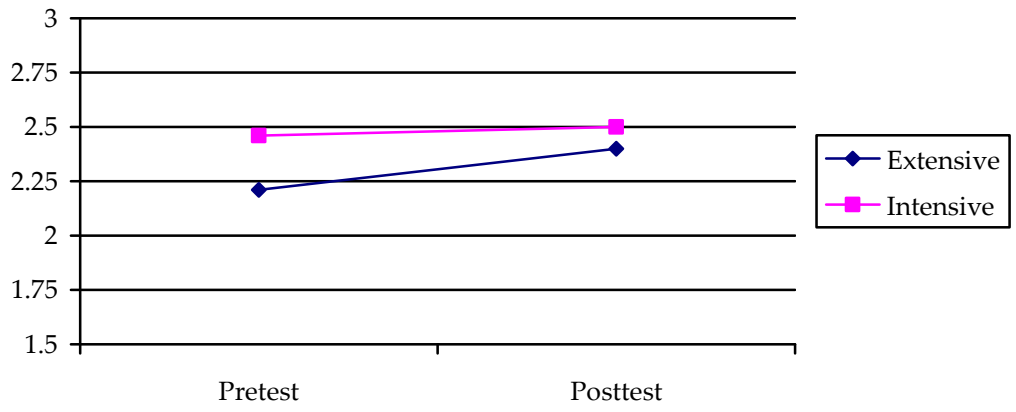


The analyses performed revealed no significant main effect for time in the **sentence conversion exercise** ($F(1,63) = .665$, $p = .418$, partial $\eta^2 = .010$). Similarly, no significant main effect was obtained for program ($F(1,63) = .396$, $p = .531$, partial $\eta^2 = .006$), or program by time ($F(1,63) = .283$, $p = .597$, partial $\eta^2 = .004$).

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there were no differences between the extensive and intensive groups in the pretest ($F(1,63) = .665$, $p = .418$, partial $\eta^2 = .010$), or in the posttest ($F(1,63) = .103$, $p = .750$, partial $\eta^2 = .002$).

Pairwise comparisons applying the Bonferroni adjustment demonstrated that the students in the two programs examined (extensive and intensive) did not make significant progress from pre to posttest in the sentence conversion exercise ($F(1,63) = .952$, $p = .333$, partial $\eta^2 = .015$; $F(1,63) = .039$, $p = .845$, partial $\eta^2 = .001$ respectively). See Figure 6.19.

Figure 6.19: Sentence Conversion Level 5

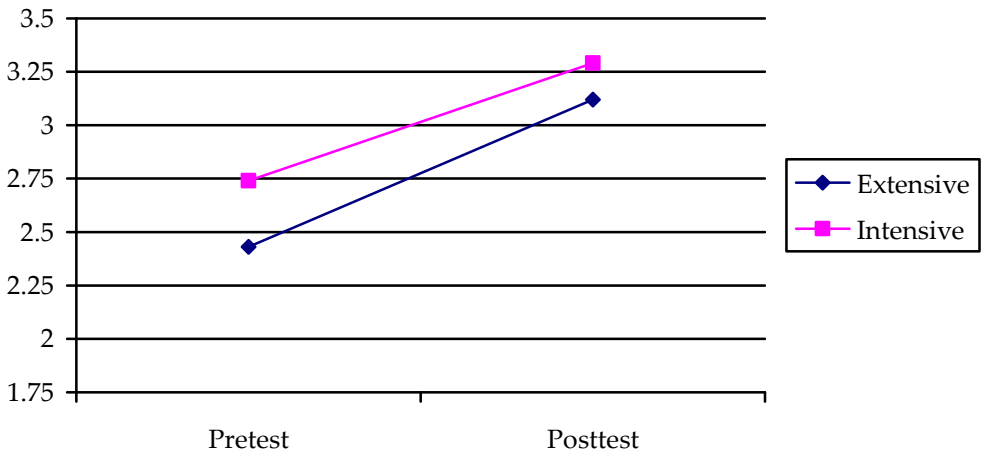


The two-way mixed ANOVA performed with the scores on **the cloze test** as the dependent variable revealed a significant main effect for time ($F(1,66) = 32.88$, $p = .000$, partial $\eta^2 = .333$), but there was no significant main effect for program ($F(1,66) = 1.26$, $p = .266$, partial $\eta^2 = .019$). Neither was the interaction effect for program by time significant ($F(1,66) = .426$, $p = .516$, partial $\eta^2 = .006$).

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there were no differences between the two programs in the pretest ($F(1,66) = 1.35$, $p = .250$, partial $\eta^2 = .020$), or in the posttest ($F(1,66) = .659$, $p = .420$, partial $\eta^2 = .010$).

Pairwise comparisons applying the Bonferroni adjustment showed that the students enrolled in both the extensive and intensive program made significant progress from the pretest to the posttest ($F(1,66) = 22.37$, $p = .000$, partial $\eta^2 = .253$; $F(1,66) = 11.87$, $p = .001$, partial $\eta^2 = .152$; respectively). See Figure 6.20.

Figure 6.20: Cloze Test Level 5

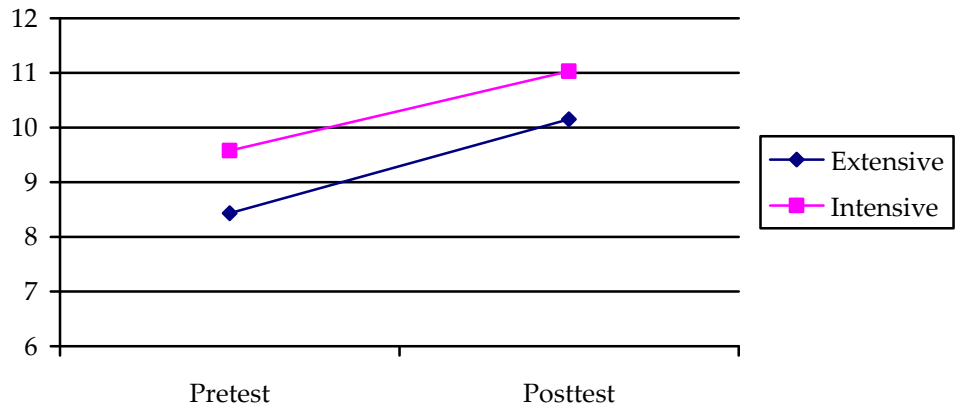


A two-way mixed ANOVA with the **total scores on the written test** as the dependent variable revealed a significant main effect for time ($F(1,63) = 32.42$, $p = .000$, partial $\eta^2 = .340$). Nevertheless, no significant main effects for program ($F(1,63) = 2.94$, $p = .091$, partial $\eta^2 = .045$), or for program by time ($F(1,63) = .233$, $p = .631$, partial $\eta^2 = .004$) were found.

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, demonstrated that there were no differences between the two groups in the pretest ($F(1,63) = 2.99$, $p = .089$, partial $\eta^2 = .045$) or in the posttest ($F(1,63) = 1.87$, $p = .176$, partial $\eta^2 = .029$).

Pairwise comparisons were run, again applying the Bonferroni adjustment. Contrasts revealed that the students in the two program types made significant progress from pretest to posttest ($F(1,63) = 20.00$, $p = .000$, partial $\eta^2 = .241$ (extensive); and $F(1,63) = 12.98$, $p = .001$, partial $\eta^2 = .171$, for the intensive program) when considering the total score on the written test. See Figure 6.21.

Figure 6.21: Total Written Test Level 5



Summarizing, as the results in Table 6.8 indicate, and as it can be clearly seen in Figure 6.18 through Figure 6.21, the students in the two programs progressed in a surprisingly similar way. As opposed to what was reported for level 3, no significant differences were found between the students in the extensive and the intensive courses (as indicated by the *p* value in the columns *program*, *program by time*, *pretest* and *posttest*). Moreover, when analyzing the variable *time*, it can be observed that, when there is progress in time, it is demonstrated by both the extensive and the intensive program. Only by observing the partial eta square can slight differences in the size of the effect be found with respect to changes in time for the students in the extensive and intensive programs in the listening and in the cloze test. Regarding the former, there is a slight larger effect size for the intensive, and in the latter for the extensive.

Consequently, it can be claimed that there are no significant interaction effects that relate improvement to intensity for the students with advanced proficiency in English, as opposed to what was presented in section 6.1.1 for the

students at the intermediate level. The level 5 students' L2 gains in the listening, sentence conversion and cloze tasks were similar, whether the type of program they followed was extensive or intensive.

Table 6.8: Results Two-Way Mixed ANOVA Written Test Level 5

	Time	Program	Program x Time	Pretest	Posttest	Pairwise comparisons	
						Extensive	Intensive
Listening	F(1,66) = 26.53 ***p*= .000** partial η^2 = .264	F(1,66) = 2.15 *p* = .147 partial η^2 = .032	*F*(1,66) = .065, *p* =.800 partial η^2 = .001	F(1,66) = 1.38 *P* =.244 partial η^2 = .020	F(1,66)= 2.01 *p*= .160 partial η^2 = .030	F(1,66)= 11.64 ***p* = .001** partial η^2 = .150	F(1,66)= 12.04 ***p* = .001** partial η^2 = .154
Sentence conversion	F(1,63) = .665 *p* = .418 partial η^2 = .010	F(1,63) = .396 *p* = .531 partial η^2 = .006	F(1,63) = .283 *p* =.597 partial η^2 = .004	F(1,63) = .665 *p* = .418 partial η^2 = .010	F(1,63) = .103 *p* = .750 partial η^2 = .002	F(1,63) = .952 *p* = .333 partial η^2 = .015	F(1,63) = .039 *p* = .845 partial η^2 = .001
Cloze	F(1,66) = 32.88 ***p* = .000** partial η^2 = .333	F(1,66) = 1.26 *p* =.266 partial η^2 = .019	F(1,66) = .426 *p* = .516 partial η^2 = .006	F(1,66) = 1.35 *p* = .250 partial η^2 = .020	F(1,66) = .659 *p* = .420 partial η^2 = .010	F(1,66)= 22.37 ***p* = .000** partial η^2 = .253	F(1,66)= 11.87 ***p* = .001** partial η^2 = .152
Total	F(1,63) = 32.43 ***p* = .000** partial η^2 = .340	F(1,63) = 2.94, *p* = .091 partial η^2 = .045	F(1,63) = .233 *p* = .631 partial η^2 = .004	F(1,63) = 2.99 *p* = .089 partial η^2 = .045	F(1,63)= 1.87 *p* = .176 partial η^2 = .029	F(1,63)= 20.00 ***p* =.000** partial η^2 = .241	F(1,63)= 12.98 ***p* = .001** partial η^2 = .171

6.2.2. Students' progress in writing skills

The students' writing samples were analyzed in terms of fluency (words per T-unit, W/T), syntactic complexity (clauses per T-unit, C/T), lexical richness (Guiraud's Index) and accuracy (percentage of error-free T-units per T-unit, EFTU/T; and errors per T-unit, Err/T). The means (M), standard deviations (SD), skewness (Sk), kurtosis (Kur) are presented in Table 6.9. Even if some measures failed the Kolmogorov-Smirnov test of normality the distribution can be said to be normal according to the results obtained when dividing the skewness and kurtosis by their standard errors in all the measures for both groups, except for the second accuracy measure (errors per T-unit), which was not normally distributed in the case of the posttest for the intensive group. Additionally, the error variance for all the measures is equal across groups, according to the Levene's tests (see Appendix L).

Table 6.9: Descriptive Statistics Writing Level 5

		EXTENSIVE				INTENSIVE			
		M	**SD**	**Sk**[40]	**Kur**	**M**	**SD**	**Sk**	**Kur**
PRETEST	**Fluency (W/T)**	11.31	3.03	.603	1.03	13.12	3.67	.496	-.787
	Syntactic C. (C/T)	1.98	.512	1.02	1.55	2.19	.671	1.19	1.42
	Lexical C. (Guiraud's I.)	7.69	.735	-.585	-.673	7.82	.607	.198	-.115
	Accuracy 1 (EFTU/T)	54.50	18.02	-.056	.160	49.48	20.33	-.293	-.006
	Accuracy 2 (Err/T)	.678	.353	.501	.278	.836	.431	.260	.041
POSTTEST	**Fluency (W/T)**	13.56	3.56	1.02	.932	14.22	3.24	.305	.211
	Syntactic C. (C/T)	2.47	.639	1.44	2.45	2.43	.566	1.09	2.27
	Lexical C. (Guiraud's I.)	7.46	.719	.285	.359	7.60	.652	-.210	-.966
	Accuracy 1 (EFTU/T)	55.09	19.19	.089	-.906	48.11	19.52	.136	-1.15
	Accuracy 2 (Err/T)	.732	.478	.792	-.290	.815	.486	1.50	3.05

As for the previous tasks, two-way mixed ANOVAs were performed in order to determine whether there were differences between the two programs in the writing measures, as well as within the same program between the two time points (pre and posttest). Writing samples were collected from 63 students, 32 of them were registered in the extensive group and 31 in the intensive.

A two-way mixed ANOVA with the **words per T-unit** measure as the dependent variable revealed a significant main effect for time ($F(1,61) = 9.14$, $p = .004$, partial $\eta^2 = .130$). There was a marginally significant main effect for program ($F(1,61) = 3.62$, $p = .062$, partial $\eta^2 = .056$), and there was no significant effect for

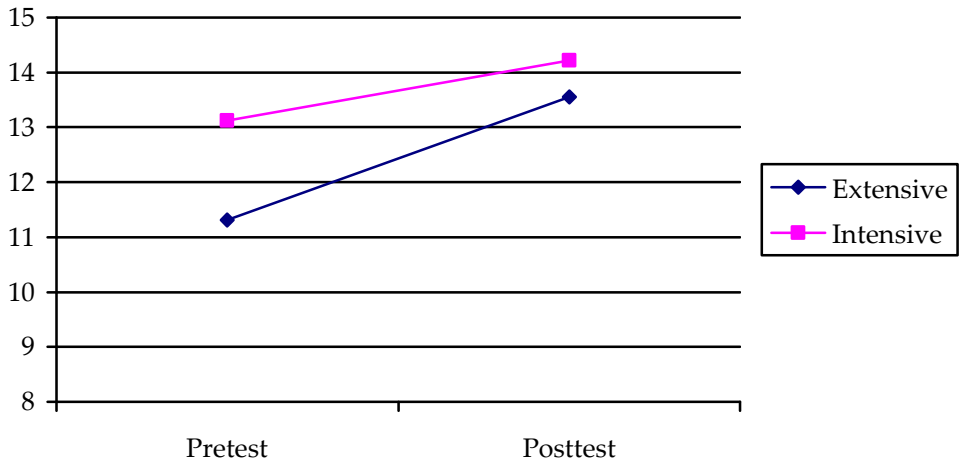
[40] For the extensive group the standard error for the skewness is .414 and for the kurtosis .809; for the intensive group .421 for the skewness and .821 for the kurtosis.

program by time ($F(1,66) = 1.09$, $p = .299$, partial $\eta^2 = .018$).

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there was a significant difference between the two groups in the pretest ($F(1,61) = 4.59$, $p = .036$, partial $\eta^2 = .070$) in favor of the students in the intensive program, yet no such difference was found in the posttest ($F(1,61) = .587$, $p = .447$, partial $\eta^2 = .010$).

Contrasts of the Pairwise comparisons, applying the Bonferroni adjustment, revealed that the students in the extensive program significantly increased the number of words per T-unit from pretest to posttest ($F(1,61) = 8.42$, $p = .005$, partial $\eta^2 = .121$), while the intensive group made only very modest gains regarding this measure from pretest to posttest, which were not statistically significant, ($F(1,61) = 1.92$, $p = .170$, partial $\eta^2 = .031$). See Figure 6.22.

Figure 6.22: Fluency (W/T) Writing L5



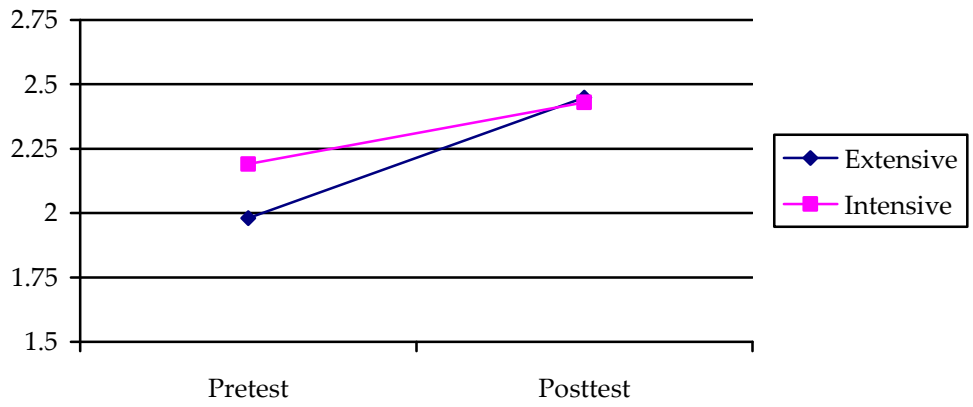
The analyses of the **clauses per T-unit** revealed a significant main effect for

time ($F(1,61) = 14.04$, $p = .000$, partial $\eta^2 = .187$). There was no significant main effect for program ($F(1,61) = .505$, $p = .480$, partial $\eta^2 = .008$), and no significant interaction effect was found for program by time ($F(1,61) = 1.87$, $p = .176$, partial $\eta^2 = .030$).

An investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there were no differences between the two program types in the pretest ($F(1,61) = 2.06$, $p = .156$, partial $\eta^2 = .033$) or in the posttest ($F(1,61) = .112$, $p = .738$, partial $\eta^2 = .002$).

Pairwise comparisons applying the Bonferroni adjustment revealed again that only the students in the extensive program had used significantly more clauses per T-unit in the posttest than in the pretest ($F(1,61) = 13.30$, $p = .001$, partial $\eta^2 = .179$), whereas the students in the intensive course also used more, yet the difference between pre and posttest was not statistically significant ($F(1,61) = 2.79$, $p = .100$, partial $\eta^2 = .044$). See Figure 6.23.

Figure 6.23: Syntactic Complexity (C/T) Writing Level 5

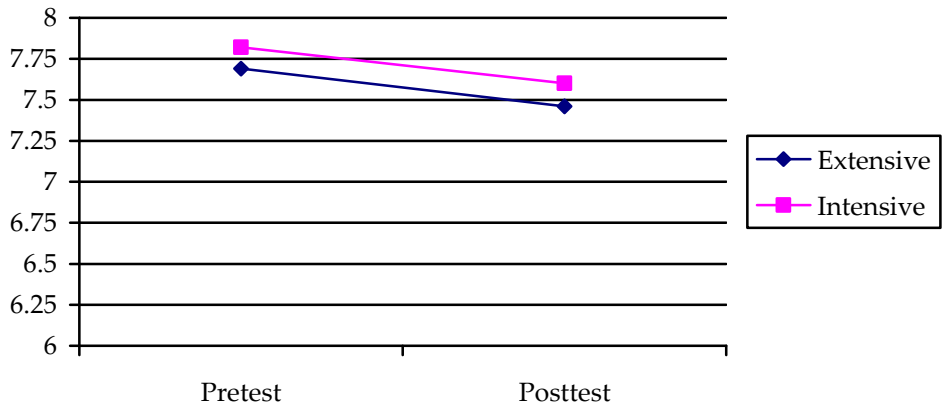


The ANOVA performed with the **Guiraud's Index of Lexical Richness** as the dependent variable revealed a significant main effect for time ($F(1,61) = 6.08$, $p = .017$, partial $\eta^2 = .091$), but there was no significant main effect for program ($F(1,61) = .839$, $p = .363$, partial $\eta^2 = .014$). Neither was the interaction effect for program by time significant ($F(1,61) = .006$, $p = .937$, partial $\eta^2 = .000$).

The investigation of the simple main effects, applying the Bonferroni adjustment for multiple comparison, showed that there were no differences between the two groups in the pretest ($F(1,61) = .545$, $p = .463$, partial $\eta^2 = .009$) or in the posttest ($F(1,61) = .656$, $p = .421$, partial $\eta^2 = .011$).

Pairwise comparisons applying the Bonferroni adjustment confirmed that there were no differences between the two groups neither in the pretest nor in the posttest. Contrasts revealed that none of the groups had made significant progress from pre to posttest in terms of lexical complexity (extensive: $F(1,61) = 3.29$, $p = .075$, partial $\eta^2 = .051$, and intensive: $F(1,61) = 2.80$, $p = .099$, partial $\eta^2 = .044$). See Figure 6.24.

Figure 6.24: Lexical Complexity (G.I.) Writing Level 5



The two-way mixed ANOVA performed with the **accuracy measures** (percentage of error-free T-units per T-unit, and errors per T-unit) revealed that there were no significant differences with respect to time, program, or program by time. The same result was found for the Pairwise comparisons applying the Bonferroni adjustments (see Table 6.10 for the *F*, *p* and partial η^2 values). Figure 6.25 and Figure 6.26 represent the students' performance with respect to the two accuracy measures under analysis.

Figure 6.25: Accuracy (EFTU/T) Writing Level 5

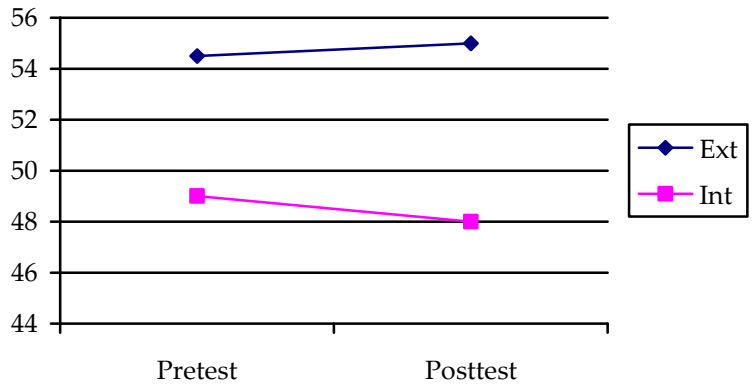


Figure 6.26: Accuracy (Err/T) Writing Level 5

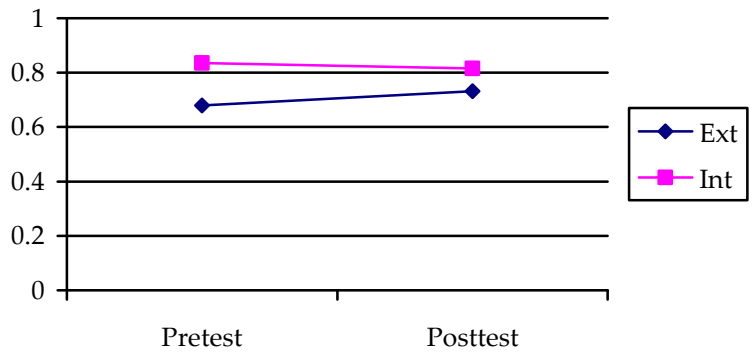


Table 6.10: Results Two-Way Mixed ANOVA Writing Level 5

	Time	Program	Program x Time	Pretest	Posttest	Pairwise comparisons	
						Extensive	Intensive
Fluency (W/T)	F(1,61) = 9.14 ***p*= .004** partial η^2 = .130	F(1,61) = 3.62 *p* = .062 partial η^2 = .056	F(1,61)=. 1.09 *p* = .299 partial η^2 = .018	F(1,61)= 4.59, ***p* = .036** partial η^2 = .070	F(1,61)= .587 *p* =.447 partial η^2 = .010	F(1,61) = 8.42 ***p* = .005** partial η^2 = .121	F(1,61) = 1.92 *p* = .170 partial η^2 = .031
Syntactic Complexity (C/T)	F(1,61) = 14.04 ***p* = .000** partial η^2 = .187	F(1,61) = .505 *p* = .480 partial η^2 = .008	F(1,61) = 1.87 *p* =.176 partial η^2 = .030	F(1,61) = 2.06 *p* = .156 partial η^2 = .033	F(1,61) = .112 *p* = .738 partial η^2 = .002	F(1,61) = 13.30 ***p* = .001** partial η^2 = .179	F(1,61) = 2.79 *p* = .100 partial η^2 = .044
Lexical Complexity (Guiraud's I)	F(1,61) = 6.08 ***p* = .017** partial η^2 = .091	F(1,61) = .839 *p* =.363 partial η^2 = .014	F(1,61) = .006 *p* = .937 partial η^2 = .000	F(1,61) = .545 *p* = .463 partial η^2 = .009	F(1,61) = .656 *p* = .421 partial η^2 = .011	F(1,61) = 3.29 *p* = .075 partial η^2 = .051	F(1,61) = 2.80 *p* = .099 partial η^2 = .044
Accuracy 1 (EFTU/T)	F(1,61) = .020 *p* = .888 partial η^2 = .000	F(1,61) = 2.26 *p* = .138 partial η^2 = .036	F(1,61) = .125 *p* = .725 partial η^2 = .002	F(1,61) = 1.07 *p* = .303 partial η^2 = .017	F(1,61) = 2.05 *p* = .158 partial η^2 = .032	F(1,61) = .023 *p* =.880 partial η^2 = .000	F(1,61) = .121 *p* = .729 partial η^2 = .002
Accuracy 2 (Err/T)	F(1,61) = .070 *p* = .792 partial η^2 = .001	F(1,61) = 1.73 *p* = .194 partial η^2 = .028	F(1,61)= .354 *p* = .554 partial η^2 = .006	F(1,61)= 2.53 *p* = .117 partial η^2 = .040	F(1,61) = .472 *p* = .495 partial η^2 = .008	F(1,61) = .376 *p* =.542 partial η^2 = .006	F(1,61) = .054 *p* = .818 partial η^2 = .001

In general, it can be said that the two programs under consideration are quite similar at the advanced level in terms of written production, as can be seen by the non-significant *p* values in the columns *program* and *program by time* in Table 6.10. When analyzing the variable *time* the most significant progress from pre to posttest experienced by the students at the advanced level in writing is reflected in fluency and complexity (especially syntactic). Such progress was significant in the case of students registered in the extensive group, but not for those in the intensive. Consequently, even if it cannot be claimed that the progress experienced in terms of writing fluency and complexity was related to intensity (there were never significant effects in *program* or *program by time*, as presented in Table 6.10), a clear distinction was revealed between the two types of program regarding L2 gains in those areas. The students in the extensive program showed more progress from pre to posttest in writing fluency and syntactic complexity than those registered in the intensive program. The results obtained for the other measures (lexical complexity and accuracy) by the students in extensive and intensive courses in highly comparable.

6.2.3. Students' progress in speaking skills

As was the case for the level 3 learners, the level 5 students' oral production was measured in terms of fluency (syllables per minute), complexity in syntax (clauses per T-unit) and in vocabulary (Guiraud's Index of Lexical Richness), and accuracy (percentage of error-free T-units per T-unit, and errors per T-unit). A

subset of the students in level 5 was randomly chosen to perform the oral tasks, which included 23 participants in the extensive program and 22 in the intensive (see Table 5.3). The results of the normality test are presented in Appendix M.

In order to examine differences between the two program types several non-parametric tests were executed due to the fact that the *n* was lower than 30 in both groups. Mann-Whitney *U* Tests were performed with the extensive and intensive groups as independent variables and each of the measures in the pre and posttest as dependent variables. The Wilcoxon Signed Rank test was employed in order to examine the effect of the variable *time*, thus analyzing the differences in the students' performance for each of the measures of oral production considered between time 1 (pretest) and time 2 (posttest). Through this test it was possible to determine whether the progress from the pretest to the posttest was significant in the case of the learners registered in the extensive and the intensive programs. Before reporting the results of the statistical analyses, the descriptive statistics including the means (M), standard deviations (SD), skewness (Sk) and kurtosis (Kur) are presented in Table 6.11.

Table 6.11: Descriptive Statistics Speaking Level 5

		EXTENSIVE				INTENSIVE			
		M	**SD**	**Sk**	**Kur**	**M**	**SD**	**Sk**	**Kur**
PRETEST	**Fluency (Syll/min)**	108.8	22.42	-.365	-.212	118.1	21.91	-.033	-1.28
	Syntactic C (C/T)	2.01	.550	.942	.105	1.83	.365	1.41	2.07
	Lexical C (Guiraud's I)	5.63	.506	-.207	-.753	5.80	.674	.212	-.801
	Accuracy 1 (EFTU/T)	52.28	16.33	-.252	-.838	49.03	17.26	-.309	-.474
	Accuracy 2 (Err/T)	.666	.297	.919	1.23	.802	.434	.692	-.438
POSTTEST	**Fluency (Syll/min)**	117.2	27.03	.431	-.628	131.7	28.73	1.24	1.44
	Syntactic C (C/T)	1.90	.334	.341	.052	1.89	.484	2.58	8.48
	Lexical C (Guiraud's I)	5.54	.552	.200	.159	5.90	.666	.275	-.604
	Accuracy 1 (EFTU/T)	57.47	17.39	-.677	.113	53.16	23.20	-.259	-.290
	Accuracy 2 (Err/T)	.552	.270	.781	-.107	.657	.411	.917	.969

When analyzing the differences between the students in the extensive and the intensive program using the Mann-Whitney *U* Test, it became apparent that there were no differences between the two groups in the pretest in any of the oral production measures, as seen in Table 6.12.

Table 6.12: Results Mann-Whitney Speaking Pretest Level 5

	U	*Z*	*p*
Pre Syll/min	195.5	-1.31	.192
Pre C/T	207.5	-1.03	.301
Pre Guiraud	219	-.772	.440
Pre EFTU/T	227.5	-.580	.562
Pre Err/T	211	-.954	.340

The same test was performed for all the measures in the posttest, where again no significant differences were found (Table 6.13). Only the results of the

lexical complexity measure (Guiraud's Index) are close to statistical significance in the posttest in favor of the students in the intensive group. Figure 6.29 illustrates that the students in the extensive group demonstrated slightly less lexical complexity in the posttest than in the pretest, while those in the intensive program showed the opposite pattern. Surprisingly, no significant differences were found in syntactic complexity, despite the fact that, as Figure 6.28 shows, the two groups progressed in different ways, with the students in the extensive program using fewer clauses per T-unit in the posttest than in the pretest, and with the learners in the intensive program following the opposite trend. As can be observed in Figure 6.27, Figure 6.30 and Figure 6.31, the students in extensive and intensive classes progressed in similar ways in terms of fluency and accuracy.

Table 6.13: Results Mann-Whitney Speaking Posttest Level 5

	U	*Z*	*p*
Post Syll/minB	187.5	-1.49	.137
Post C/T	219	-.773	.440
Post Guiraud	177	-1.73	.084
Post EFTU/T	228.5	-.557	.578
Post Err/T	223.5	-.670	.503

With the purpose of examining the effect of *time* in the performance of the students in the two program types, Wilcoxon Signed Rank Tests were executed. The results of the test indicated that the only measure where the students in the **extensive** program made significant progress was fluency (Z -2.25, p = .024). No significant improvement was found for syntactic complexity (Z -1.04, p = .299), lexical complexity (Z -.365, p = .715) or accuracy (EFTU/T: Z -1.22, p = .223; Err/T: Z -1.51, p = .131).

Interestingly, similar results were found in the same test for the students in the **intensive** group. Significant progress was made by these learners in fluency (*Z* -2.48, *p* = .013). However, no significant difference was observed in terms of syntactic complexity (*Z* -.435, *p* = .664), lexical complexity (*Z* -1.12, *p* = .263), or accuracy (EFTU/T: *Z* -.568, *p* = .570; Err/T: *Z* -1.48, *p* = .140).

In sum, no significant interaction effects that related improvement to intensity were found, since, first of all, there are no differences between the two program types in the pretest or in the posttest in any of the measures chosen to analyze oral production. Furthermore, the progress made by the students in the extensive and intensive programs is highly comparable in terms of fluency and accuracy, with the students in the two programs improving significantly in terms of the former and just showing some minor improvement with respect to the latter. However, opposite trends between the two program types are found regarding syntactic and lexical complexity. Although the scores in the clauses per T-unit measure and in the Guiraud's Index are not significantly different for the students in the extensive and the intensive course, and no significant improvement is experienced by any of the groups from pretest to posttest, it must be noticed that the students in the intensive group made gains in syntactic and lexical complexity (despite them not being significant), whereas the students in the extensive program obtained lower scores in the measures considered for those language areas in the posttest than in the pretest.

Figure 6.27: Fluency (Syll/Min) Speaking Level 5

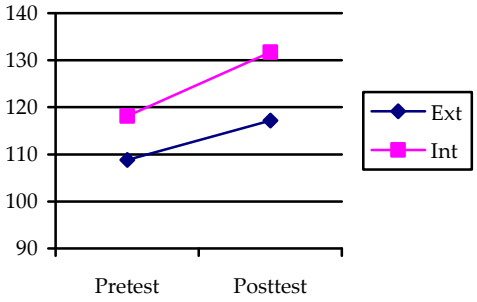


Figure 6.28: Synt. C. (C/T) Sp Level 5

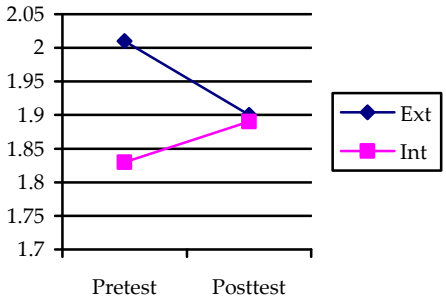


Figure 6.29: Lex. C. (G.I.) Sp Level 5

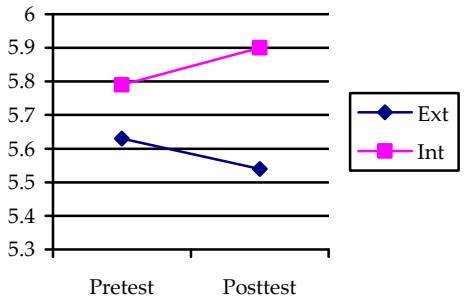


Figure 6.30: Acc. (EFTU/T) Sp Level 5

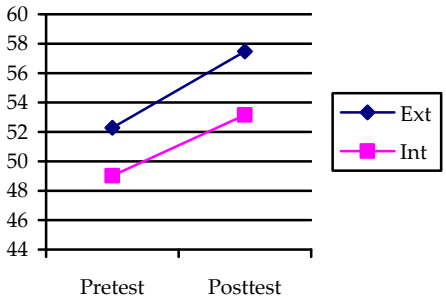
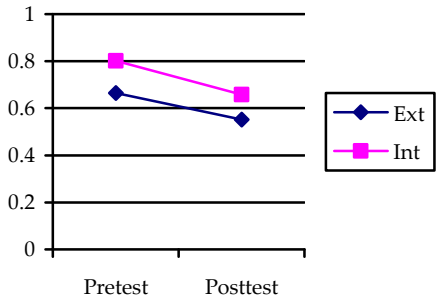


Figure 6.31: Acc. (Err/T) Sp Level 5



6.3. Summary of the results level 3 and level 5

Table 6.14 provides a summary of the results for the intermediate and advanced students, showing where statistically significant differences were found with respect to the two main variables under consideration: program type and time. A statistically significant result is indicated with an asterisk (*) on top of the name of the group with the significantly higher score (in comparisons) or on top of the group whose scores were significantly different from pre to posttest (column *time*[41]).

[41] If there are no other specifications, the result in the posttest is supposed to be higher than in the pretest.

Table 6.14: Summary Results Level 3 & Level 5 All Measures

	Measures	Program		Time
		Pretest	**Posttest**	
WRITTEN TEST	**Listening**		L3 (S*-E, I*-E)	L3 (E*, S*, I*) *L5 (E*, I*)*
	Sent Conversion		L3 (I*-E)	L3 (E*, S*, I*)
	Cloze			L3 (E*, S*, I*) *L5 (E*, I*)*
	Reading Total			L3 (E*, S*, I*)
	TOTAL test		L3 (S*-E, I*-E)	L3 (E*, S*, I*) *L5 (E*, I*)*
WRITING	**Fluency (W/T)**	*L5 (I*-E)*	L3 (E*-S)	L3 (E*, I*) *L5 (E*)*
	Syntactic Complexity (C/T)		L3 (E*-S, I*-S)	L3 (E*, I*) *L5 (E*)*
	Lexical Compl (Guiraud's I)		L3 (I*-E, I*-S)	L3 (E*, S* worse; I* better)
	Accuracy 1 (EFTU/T)		L3 (S*-E)	L3 (E* worse)
	Accuracy 2 (Err/T)		L3 (S*-E)	L3 (E* worse)
SPEAKING	**Fluency (syll/min)**	L3 (S*-E, I*-E)		L3 (E*, I*) *L5 (E*, I*)*
	Syntactic Complexity (C/T)	L3 (I*-S)		L3 (E*, S*)
	Lexical Compl (Guiraud's I)			L3 (I*)
	Accuracy 1 (EFTU/T)			L3 (I*)
	Accuracy 2 (Err/T)			L3 (I*)

L3: Level 3 (interm) L5: Level 5 (adv) E: Extensive, S: Semi-intensive, I: Intensive

It can be observed that in the first five measures, which refer to the exercises in the **written test** that all the students took, there is a clear effect of intensity in level 3, as reflected in the *program* variable. The students in the semi-intensive program had significantly superior scores in the posttest on the listening exercise than their peers in the extensive group. Moreover, the learners in the intensive program had significantly higher scores than those in the extensive course on both the listening and the sentence conversion exercises. In terms of the students´ performance in the cloze activity (which evaluates grammar skills) and the students´ reading skills, an interaction effect could not be found between intensity and the students' performance. With respect to the *time* variable, statistically significant effects were found for the three program types at this level for all the exercises included in the written test (listening, sentence conversion, cloze, and reading). At level 5 there are no interaction effects that relate improvement to intensity, as observed in the students' performance in the listening, sentence conversion or cloze tasks. No significant differences were found in the *program* or *time* variables for the advanced students in the extensive and intensive programs.

With respect to the **writing measures**, the results obtained in level 3 taking into account the *program* variable show a certain advantage for the students in the extensive and intensive programs in terms of fluency and syntactic complexity with respect to those in the semi-intensive group. Conversely, the students in the latter group have a more accurate writing production. Finally, the vocabulary used by the students in the intensive program in the posttest seems to be more lexically

complex than the vocabulary produced by the learners in the other two programs. Then, analyzing the variable *time*, a certain advantage can be attributed to the students in the level 3 intensive group, since they made progress in fluency, syntactic and lexical complexity without such progress having a negative effect on accuracy, as happened for those registered in the extensive group, who made progress in fluency and syntactic complexity but demonstrated a significantly less accurate writing production in the posttest. The students in the semi-intensive group did not show any significant progress. The case of the level 5 students is quite different, since intensity is not playing an important role, yet if some students can be said to have improved more clearly their writing skills, it would be those in the extensive program, as reflected in the significant progress in fluency and syntactic complexity when analyzing *time*.

Regarding **oral production**, in the case of the intermediate-level students, again, a slight[42] advantage should be attributed to the learners in the intensive program, since they significantly progressed in all the measures analyzed except for syntactic complexity, while the students in the extensive program only showed significant progress in fluency and syntactic complexity (see column time Table 6.14). Again, the students registered in the semi-intensive group did not show much improvement, their progress only being significant in the syntactic complexity measure. The students in level 5 demonstrated highly similar speaking skills in the two program types, making significant progress in the same measure

[42] The advantage cannot be said to be very pronounced because, as reflected in Table 6.14 there were no significant differences in the posttest across programs.

(fluency).

6.4. Comparisons between level 3 and level 5 students in written and oral production

The results obtained in all the measures which were used to evaluate written and oral production for the level 3 students were compared with the results achieved by the level 5 learners in order to examine whether such measures were adequate to analyze progress from an intermediate to an advanced proficiency level[43]. It can be assumed that the differences obtained in specific measures among programs of the same level should be reliable if the same measures capture differences between different levels of proficiency, which are, in principle, more obvious. Section 6.4.1 will present the comparisons for the writing measures, whereas section 6.4.2 will be concerned about the students' oral production.

[43] Apart from the comparisons which will be presented here (which deal with the level 3 students as a group and the level 5 students), comparisons were established between each of the level 3 programs (extensive, semi-intensive and intensive) and the level 5 as a whole, in order to make sure that the level 5 students' performance differed from each of the level 3 programs in similar ways. It was found that the extensive, semi-intensive and intensive level 3 students differed significantly in the same areas as those reported in Table 6.16, except for the semi-intensive level 3 students, who were not significantly different in the pretest from the advanced students in terms of lexical complexity.

6.4.1. Written production

Table 6.15 shows the descriptive statistics for all the writing measures in the case of the intermediate (level 3) students and the advanced (level 5) students. There were 103 students in the former group and 63 in the latter. Appendix N includes the results of the normality and Levene's tests for each of the measures.

Table 6.15: Descriptive Statistics Writing Level 3 and Level 5

		INTERMEDIATE (L3)				ADVANCED (L5)			
		Mn	**SD**	**Sk**[44]	**Kur**[45]	**Mn**	**SD**	**Sk**	**Kur**
PRETEST	**Fluency (W/T)**	8.6	1.4	.17	-.60	12.2	3.4	.62	-.05
	Synt. C. (C/T)	1.5	.26	.66	.60	2.1	.60	1.2	1.8
	Lexical C. (Guiraud´s)	7.3	.77	.14	-.33	7.7	.67	-.35	-.28
	Accuracy 1 (EFTU/T)	54.9	17.4	-.14	-.71	52	19.2	-.23	.08
	Accuracy 2 (Err/T)	.68	.37	.81	-.00	.76	.40	.44	.12
POSTTEST	**Fluency (W/T)**	10.2	2.1	.75	.35	13.9	3.4	.66	.34
	Synt. C. (C/T)	1.8	.40	.75	.52	2.4	.60	1.3	2.2
	Lexical C. (Guiraud´s)	6.9	.82	-.07	-.61	7.5	.68	.04	-.32
	Accuracy 1 (EFTU/T)	50.1	18.1	.23	-.44	51.6	19.5	.10	-.99
	Accuracy 2 (Err/T)	.80	.40	.46	-.06	.77	.48	1.1	1.3

Comparing the performance in written production for the level 3 students

[44] The standard error for the skewness is .238 for the intermediate students and .302 for the advanced.

[45] The standard error for the kurtosis is .472 for the intermediate students and .595 for the advanced.

as a whole with all the level 5 students, it was evident that the two levels were significantly different from each other in all the areas, according to the independent measures T-test (Table 6.16) performed, except for accuracy.

Table 6.16: Results T-test Writing Level 3 vs. Level 5

		t	*df*	*p*
PRETEST	**Fluency (W/T)**	-7.93	75.17	**.000**
	Synt. C. (C/T)	-7.06	76.39	**.000**
	Lexical C. (Guiraud´s)	-4.03	164	**.000**
	Accuracy 1 (EFTU/T)	1.01	164	.313
	Accuracy 2 (Err/T)	.-503	182	.616
POSTTEST	**Fluency (W/T)**	-7.73	90.99	**.000**
	Synt. C. (C/T)	-7.31	96.20	**.000**
	Lexical C. (Guiraud´s)	-4.47	164	**.000**
	Accuracy 1 (EFTU/T)	-.506	164	.613
	Accuracy 2 (Err/T)	.949	182	.344

The case of accuracy in written production was different from the other measures in terms of expected outcomes from level 3 and level 5 students, with respect to the fact that neither in the pretest nor in the posttest were the level 5 students more accurate (as reflected in the production of more error-free T-units per T-unit or fewer errors per T-unit) than their level 3 peers.

6.4.2. Oral production

The descriptive statistics for the scores on all the measures of oral production for the level 3 and the level 5 students are reported in Table 6.17. The number of students included in the intermediate group was 53, while those

registered in the advanced were 45.

Table 6.17: Descriptive Statistics Speaking Level 3 & Level 5

		INTERMEDIATE (L3)				ADVANCED (L5)			
		Mn	**SD**	**Sk**[46]	**Kur**[47]	**Mn**	**SD**	**Sk**	**Kur**
PRETEST	**Fluency (Syll/Min)**	71.1	19.3	.23	.-06	113.4	22.4	-.20	-.55
	Synt. C. (C/T)	1.53	.37	1.07	.92	1.9	.47	1.2	1.02
	Lexical C. (Guiraud´s)	4.9	.54	-.09	-.99	5.7	.59	.20	-.49
	Accuracy 1 (EFTU/T)	30.9	19.1	.09	-.53	50.7	16.7	-.28	-.66
	Accuracy 2 (Err/T)	1.12	.53	.99	.82	.73	.37	.92	.37
POSTTEST	**Fluency (Syll/Min)**	80.0	17.6	.28	-.59	124.3	28.5	.798	.753
	Synt. C. (C/T)	1.64	.35	.55	-.49	1.9	.41	2.0	7.0
	Lexical C. (Guiraud´s)	5.06	.52	-.27	.27	5.7	.63	.36	-.18
	Accuracy 1 (EFTU/T)	38.7	20.0	.34	-.70	55.4	20.3	-.47	-.08
	Accuracy 2 (Err/T)	1.05	.46	.91	1.5	.60	.35	1.1	1.5

The results of the normality and Levene's test of equality of error variances are presented in Appendix O. Table 6.18 shows the results of the T-test, which indicate that there were significant differences in all the measures adopted to analyze oral production in the case of the students at the intermediate and at the advanced level both in the pretest and in the posttest.

[46] The standard error for the skewness is .327 for the intermediate students and .354 for the advanced.

[47] The standard error for the kurtosis is .644 for the intermediate students and .695 for the advanced.

Table 6.18: Results T-test Speaking Level 3 vs. Level 5

		t	*df*	*p*
PRETEST	**Fluency (Syll/Min)**	-10.03	96	**.000**
	Synt. C. (C/T)	-4.59	96	**.000**
	Lexical C. (Guiraud´s)	-6.61	96	**.000**
	Accuracy 1 (EFTU/T)	-5.39	96	**.000**
	Accuracy 2 (Err/T)	4.27	93.01	**.000**
POSTTEST	**Fluency (Syll/Min)**	-9.04	70.93	**.000**
	Synt. C. (C/T)	-3.36	96	**.001**
	Lexical C. (Guiraud´s)	-5.65	96	**.000**
	Accuracy 1 (EFTU/T)	-4.07	96	**.000**
	Accuracy 2 (Err/T)	5.35	96	**.000**

6.5. Comparison between advanced learners and NESs

Written and oral data were collected from native-English speakers (NESs) in order to provide a baseline for comparison, especially with respect to the performance of the advanced learners included in this research. When the scores of the level 5 students are compared with those of native speakers in each of the measures considered for written and oral production, it should become evident whether the measures are adequate to analyze progress at the advanced level (in the case where the scores obtained by the advanced students are different from the NESs) or whether there are ceiling effects (when advanced learners do not differ from NESs). According to the information presented in sections 6.2.2 and 6.2.3, the level 5 students did not experience much improvement in the measures examining written and oral production. If differences between these learners and NESs exist, it could be argued that, despite the fact that there was room for progress, for whatever reason, these learners did not demonstrate any language gains. If, on the

other hand, the performance of the advanced students is shown to be comparable to that of NESs, it could be claimed that the students did not improve their performance because it was near native-like with respect to the measures considered.

As reported in section 6.2.2, there were significant differences between the students in the extensive and the intensive programs at level 5 in the pretest regarding written fluency. Besides, it was claimed that, while the students in the extensive program made significant progress from pre to posttest in written fluency and syntactic complexity, no such progress was observed in the case of the students in the intensive group. Consequently, it was considered appropriate to compare the performance of each of these groups separately with the performance of NESs in order to examine whether the reason why no significant progress was made by the students in the intensive group was due to the fact that their performance was already quite native-like in terms of written production in the pretest. The results of the comparisons will be presented first for written skills (section 6.5.1) and then for speaking skills (section 6.5.2).

6.5.1. Written production

The descriptive statistics for the written production measures for the level 5 extensive and intensive groups can be seen in Table 6.7, yet they will be reproduced in Table 6.19 to facilitate the comparison with the NESs.

Table 6.19: Descriptive Statistics Writing Level 5 and NESs

		EXTENSIVE				INTENSIVE				NES			
		Mean	**SD**	**Sk**	**Kur**	**Mean**	**SD**	**Sk**	**Kur**	**Mean**	**SD**	**Sk**	**Kur**
PRETEST	**Fluency (W/T)**	11.31	3.03	.603	1.03	13.12	3.67	.496	-.787	14.40	3.61	.575	-1.08
	Syntactic C. (C/T)	1.98	.512	1.02	1.55	2.19	.671	1.19	1.42	2.09	.605	1.47	1.48
	Lexical C. (Guiraud's)	7.69	.735	-.585	-.673	7.82	.607	.198	-.115	8.07	.574	-1.30	2.51
	Accuracy 1 (EFTU/T)	54.50	18.02	-.056	.160	49.48	20.33	-.293	-.006	95.66	4.27	.043	-2.17
	Accuracy 2 (Err/T)	.678	.353	.501	.278	.836	.431	.260	.041	.043	.042	-.021	-2.14
POSTTEST	**Fluency (W/T)**	13.56	3.56	1.02	.932	14.22	3.24	.305	.211	14.40	3.60	.575	-1.07
	Syntactic C. (C/T)	2.47	.639	1.44	2.45	2.43	.566	1.09	2.27	2.09	.604	1.47	1.48
	Lexical C. (Guiraud's)	7.46	.719	.285	.359	7.60	.652	-.210	-.966	8.06	.574	-1.30	2.51
	Accuracy 1 (EFTU/T)	55.09	19.19	.089	-.906	48.11	19.52	.136	-1.15	95.66	4.27	.043	-2.17
	Accuracy 2 (Err/T)	.732	.478	.792	-.290	.815	.486	1.50	3.05	.043	.042	-.021	-2.14

SD: Standard Deviation; Sk: Skewness; Kur: Kurtosis

The native English speakers did not perform a pretest and a posttest; however, the scores are repeated for the two times in order to make it easier to compare such results with the scores of the other two groups. The results of the tests of normality can be seen in Appendix P. Since the group of NESs is smaller than 30 (there were 14 students in the sample), non-parametric statistical analyses were performed in order to examine whether there were significant differences between the students at the advanced level (32 in the extensive group and 31 in the intensive) and the NESs. First, the Kruskal-Wallis test was performed so as to analyze differences between the three groups considered (extensive level 5, intensive level 5, and NESs). Then, Mann-Whitney *U* tests were executed in order to analyze differences between each of the level 5 groups and the NESs.

The Kruskal-Wallis test, including the students in the extensive level 5, intensive level 5, and the NESs indicated that there were significant differences among the three groups in terms of fluency and accuracy in the pretest, and syntactic, lexical complexity, and accuracy in the posttest (see Table 6.20).

Table 6.20: Results Kruskal-Wallis Writing Level 5 vs. NESs

		χ^2	*df*	*p*
PRETEST	**Fluency (W/T)**	6.31	2	**.043**
	Synt Compl (C/T)	1.44	2	.486
	Lexical C. (Guiraud)	2.89	2	.236
	Accuracy 1 (EFTU/T)	31.50	2	**.000**
	Accuracy 2 (Err/T)	31.20	2	**.000**
POSTTEST	**Fluency (W/T)**	1.57	2	.455
	Synt Compl (C/T)	6.30	2	**.043**
	Lexical C. (Guiraud)	9.02	2	**.011**
	Accuracy 1 (EFTU/T)	32.10	2	**.000**
	Accuracy 2 (Err/T)	31.46	2	**.000**

In section 6.2.2 the results of the statistical analyses comparing performance in writing for the students in the **extensive and intensive level 5** programs were presented. Such analyses showed that the students in the intensive program had significantly higher scores in fluency than those in the extensive in the pretest. In the posttest no differences were registered between the two level 5 groups (see Table 6.10).

In order to compare the performance of each of the level 5 groups and the NESs, Mann-Whitney *U* tests were executed. The results of such test indicated that, in the pretest, there were significant differences between the students in the **extensive level 5 program and the NESs** in fluency in terms of words per T-unit, and especially in the two accuracy measures. The written production of the extensive level 5 learners and the NESs was not significantly different in the pretest in syntactic or lexical complexity (see Table 6.21).

Table 6.21: Results Mann-Whitney *U* Test Writing Extensive Level 5 vs. NESs

		U	*Z*	*P*
PRETEST	**Fluency (W/T)**	115	-2.33	**.020**
	Synt Compl (C/T)	188.5	-.488	.625
	Lexical C. (Guiraud)	146.5	-1.54	.124
	Accuracy 1 (EFTU/T)	7.00	-5.04	**.000**
	Accuracy 2 (Err/T)	9.00	-4.99	**.000**
POSTTEST	**Fluency (W/T)**	174	-.851	.395
	Synt Compl (C/T)	112	-2.40	**.016**
	Lexical C. (Guiraud)	89	-2.98	**.003**
	Accuracy 1 (EFTU/T)	6.00	-5.07	**.000**
	Accuracy 2 (Err/T)	6.00	-5.06	**.000**

As a consequence, the two areas in which progress should be more clearly expected for the students in the extensive level 5 group would be fluency and accuracy, since their results in the other measures are already quite close to native-like performance in the pretest. The results of the Mann-Whitney *U* test performed for the posttest measures revealed that, in fact, there were no significant differences in fluency between the advanced level 5 learners in the extensive group and the NESs. On the other hand, significant differences still existed in the posttest regarding accuracy (Table 6.21).

Furthermore, contrarily to the results of the analyses in the pretest, it can be observed that there were significant differences in the posttest between these two groups in syntactic complexity and lexical richness. The former is due to the fact that the extensive level 5 students produced T-units which were more complex, syntactically speaking, than the NESs. The significant differences in the Guiraud's Index occurred because the extensive level 5 students produced less lexically complex language in the posttest than in the pretest; consequently, the NESs' vocabulary was richer.

With respect to the students in the **intensive level 5** group, no significant differences were obtained in the Mann-Whitney *U* test which compared their performance in the pretest with that of **NESs** except in the case of accuracy (see Table 6.22). In the posttest significant differences occurred in the syntactic complexity measure (C/T), with the intensive level 5 students producing more syntactically complex T-units; and in lexical complexity, due to the fact that the level 5 students demonstrated less lexical complexity in the posttest than in the

pretest.

Table 6.22: Results Man-Whitney *U* Test Writing Intensive Level 5 & NESs

		U	*Z*	*P*
PRETEST	**Fluency (W/T)**	159	-1.09	.274
	Synt Compl (C/T)	181.5	-.514	.607
	Lexical C. (Guiraud)	141	-1.56	.120
	Accuracy 1 (EFTU/T)	.000	-5.19	**.000**
	Accuracy 2 (Err/T)	7.00	-5.01	**.000**
POSTTEST	**Fluency (W/T)**	197	-.116	.908
	Synt Compl (C/T)	116	-2.20	**.028**
	Lexical C. (Guiraud)	114.5	-2.24	**.025**
	Accuracy 1 (EFTU/T)	.000	-5.19	**.000**
	Accuracy 2 (Err/T)	.000	-5.19	**.000**

In general, then, it can be said that the students in the extensive level 5 group were slightly less native-like in the pretest in terms of the measures considered to analyze written production than the students in the intensive program. In the posttest, the learners in both, the extensive and the intensive programs differ in similar terms with respect to NESs. Among all the different measures considered, accuracy is the area in which the performance of non-native English speakers is more distant from NESs' performance both in the pretest and in the posttest.

6.5.2. Oral production

The descriptive statistics for the speaking measures for the level 5 students and the NESs are presented in Table 6.23. A total of 23 students were included in

the extensive group, 22 in the intensive, and 13 in the NESs[48]. Appendix Q presents the results of the normality test.

[48] One student from this group did the writing task but not the speaking activity.

Table 6.23: Descriptive Statistics Speaking Level 5 & NESs

		EXTENSIVE				INTENSIVE				NESs			
		Mean	**SD**	**Sk**	**Kur**	**Mean**	**SD**	**Sk**	**Kur**	**Mean**	**SD**	**Sk**	**Kur**
PRETEST	**Fluency (Syll/min)**	108.8	22.42	-.365	-.212	118.1	21.91	-.033	-1.28	194.0	42.95	.088	-1.60
	Syntactic C. (C/T)	2.01	.550	.942	.105	1.83	.365	1.41	2.07	1.73	.270	-.643	1.62
	Lexical C. (Guiraud's)	5.63	.506	-.207	-.753	5.80	.674	.212	-.801	6.11	.775	.248	-1.25
	Accuracy 1 (EFTU/T)	52.28	16.33	-.252	-.838	49.03	17.26	-.309	-.474	99.40	2.06	-3.46	12.0
	Accuracy 2 (Err/T)	.666	.297	.919	1.23	.802	.434	.692	-.438	.012	.030	2.08	2.82
POSTTEST	**Fluency (Syll/min)**	117.2	27.03	.431	-.628	131.7	28.73	1.24	1.44	194.0	42.95	.088	-1.60
	Syntactic C. (C/T)	1.90	.334	.341	.052	1.89	.484	2.58	8.48	1.73	.270	.643	1.62
	Lexical C. (Guiraud's)	5.54	.552	.200	.159	5.90	.666	.275	-.604	6.11	.775	.088	-1.60
	Accuracy 1 (EFTU/T)	57.47	17.39	-.677	.113	53.16	23.20	-.259	-.290	99.40	2.06	-3.46	12.0
	Accuracy 2 (Err/T)	.552	.270	.781	-.107	.657	.411	.917	.969	.012	.042	2.08	2.82

As for written production, Kruskal-Wallis tests were performed in order to examine differences between the three groups under consideration. Then, Mann-Whitney *U* tests were executed so as to analyze results in pairs. The Kruskal-Wallis including the results of the extensive, intensive level 5 students, and NESs revealed that there were significant differences between the three groups in the pretest in terms of fluency and accuracy. In the posttest, the same areas remained significantly different for the three groups, and the results of the Guiraud´s Index were marginally significant (see Table 6.24).

Table 6.24: Results Kruskal-Wallis Speaking Level 5 vs. NESs

		χ^2	*df*	*p*
PRETEST	**Fluency (Syll/min)**	26.51	2	**.000**
	Synt Compl (C/T)	1.77	2	.414
	Lexical C. (Guiraud)	3.12	2	.210
	Accuracy 1 (EFTU/T)	28.39	2	**.000**
	Accuracy 2 (Err/T)	28.68	2	**.000**
POSTTEST	**Fluency (Syll/min)**	22.16	2	**.000**
	Synt Compl (C/T)	1.51	2	.470
	Lexical C. (Guiraud)	5.68	2	.058
	Accuracy 1 (EFTU/T)	28.36	2	**.000**
	Accuracy 2 (Err/T)	28.38	2	**.000**

The results of the statistical analyses which compared the oral performance of the learners registered in the **extensive and the intensive** advanced level can be seen in Table 6.12 and Table 6.13, in which no significant differences were reported between the two groups in the pretest or in the posttest.

The Mann-Whitney *U* test revealed significant differences in the pretest between the performance of the students in the **extensive level 5** group as

compared to the **NESs** concerning fluency, and especially accuracy, as can be observed in Table 6.25. Despite the fact that the extensive level 5 students progressed in terms of oral fluency in the posttest (see section 6.2.3), their performance at that time is still significantly different from the NESs' performance. With respect to the other measures, it can be said that the extensive level 5 students show a native-like performance in speaking in relation to syntactic complexity in both pre and posttest. As regards lexical complexity, differences between the two groups exist only in the posttest due to the fact that the level 5 students demonstrated less complex vocabulary at that time than in the pretest. As happened in written production, the extensive level 5 learners' accuracy is still quite far from being native-like even in the posttest.

Table 6.25: Results Mann-Whitney *U* Test Speaking Extensive L5 vs. NESs

		U	*Z*	*p*
PRETEST	**Fluency (Syll/min)**	3	-4.69	**.000**
	Synt Compl (C/T)	104	-1.18	.248
	Lexical C. (Guiraud)	89.0	-1.70	.092
	Accuracy 1 (EFTU/T)	.000	-4.87	**.000**
	Accuracy 2 (Err/T)	.000	-4.85	**.000**
POSTTEST	**Fluency (Syll/min)**	16	-4.24	**.000**
	Synt Compl (C/T)	102	-1.25	.221
	Lexical C. (Guiraud)	75	-2.19	**.028**
	Accuracy 1 (EFTU/T)	.000	-4.87	**.000**
	Accuracy 2 (Err/T)	.000	-4.85	**.000**

Concerning the oral production of the **intensive level 5** students in the pretest, it can be said to be comparable to that of **NESs** in all the measures except for oral fluency and accuracy. The students registered in the intensive level 5

program showed native-like performance in speaking in the other two measures (C/T and Guiraud's Index) in the pretest and in the posttest (see Table 6.26). As was the case for the extensive level 5 group, those students in the intensive program significantly improved their oral fluency in the posttest; nevertheless, their fluency at that time is still significantly different from NES's fluency. Similarly, the intensive level 5 students' accuracy in oral production is not native-like in the pretest or in the posttest.

Table 6.26: Results Mann-Whitney *U* Test Speaking Intensive L5 vs. NESs

		U	*Z*	*p*
PRETEST	**Fluency (Syll/min)**	10	-4.40	**.000**
	Synt Compl (C/T)	124	-.289	.790
	Lexical C. (Guiraud)	99	-1.19	.245
	Accuracy 1 (EFTU/T)	.000	-4.84	**.000**
	Accuracy 2 (Err/T)	.000	-4.82	**.000**
POSTTEST	**Fluency (Syll/min)**	22	-3.96	**.000**
	Synt Compl (C/T)	122	-.361	.736
	Lexical C. (Guiraud)	111	-.757	.466
	Accuracy 1 (EFTU/T)	.000	-4.84	**.000**
	Accuracy 2 (Err/T)	.000	-4.81	**.000**

It can be said that the students' speaking skills at the advanced level are quite comparable in the two program types (extensive and intensive), as opposed to what was found for writing skills (see section 6.5.1). Moreover, the students in the two programs differ in terms of the same measures with respect to NESs, namely, fluency and accuracy.

In the next chapter, all the results presented in Chapter 6 will be discussed in light of previous studies on the issue of learning and time distribution.

Moreover, the information obtained from the students' background questionnaire will be used in order to interpret the findings from this study. At the end of the chapter, a short section will be included to discuss the appropriateness of the measures chosen in this study to analyze written and oral production.

CHAPTER 7: DISCUSSION

The main questions that guide this dissertation refer to whether time distribution has an effect on students' L2 gains in an EFL course (Research Question 1) and whether such effect is different according to the learners' initial proficiency level (Research Question 2), as stated in Chapter 4. The results of the tests that the students performed, which were presented in Chapter 6, indicated that there is certainly an effect of time distribution on students' performance (as demonstrated by the significant differences in the statistical analyses depending on the type of program). Such effect, however, was more evident for some tasks than for others; furthermore, the distribution of instruction hours affected the intermediate-level learners and the advanced learners differently. In this chapter, the results of each of the tasks for each level under consideration will be discussed in light of previous research as well as information referring to classroom practice obtained from the teachers and the students involved in this study. The first section will be devoted to the intermediate (level 3) learners, whereas the second will be concerned about the students at the advanced level (level 5). At the end of the section a summary will be provided.

7.1. Level 3 (intermediate)

The results obtained by the intermediate-level students will be discussed first in terms of language gains in listening, grammar, vocabulary and reading

(which were the targets of the test which the students took) (section 7.1.1). Then, the results of the written production task will be examined (section 7.1.2). Section 7.1.3 will be devoted to analyzing the results of the oral production task. Initially, all the findings will be discussed in light of previous research in the program evaluation literature, then according to studies in cognitive psychology; finally, a relationship will be established between the results obtained for the different measures and the information obtained from both students and teachers regarding classroom practice.

7.1.1. Time distribution and performance in listening, grammar, vocabulary and reading exercises

The results of the statistical analyses of the data from this research suggest that, in the case of **listening skills** (as assessed by the measure included in this study), concentrating the hours of foreign language instruction seems to have a beneficial effect, since the students demonstrated more and more progress in this area as the concentration of instruction hours increased, as can be observed in Figure 6.1 or in Table 6.2.

That intensity of exposure enhances students' listening skills is not unexpected, as intensive language courses have usually emphasized audio-oral skills since the first half of the 20th century with the Army Specialized Training Program. Agard (1946) and Agard et al. (1945) reported on the advanced oral comprehension skills achieved by the soldiers at the end of their intensive foreign

language courses. Similarly, other types of intensive instruction, such as French immersion programs, have been known to promote high levels of oral comprehension abilities, which can be characterized as native-like (Genesee, 1994; 2004). Reports on intensive English in Canada have also shown advantages in listening comprehension for the learners in intensive English programs with respect to those students following traditional English instruction in their same grade but who had received fewer hours of instruction, as well as with respect to students who had the same hours of instruction but were at a higher grade (Lightbown & Spada, 1994).

The highly positive results in terms of progress in listening skills for students following concentrated foreign language instruction can be explained by the fact that the constant exposure to the L2 in intensive programs probably makes the learners' ears become more 'tuned' to the foreign language patterns, since they get used to making the form-meaning mappings from the oral input when they listen to it for long stretches of time, as is the case of the intensive program under analysis. The students who attend extensive courses only have the opportunity to listen to the L2 for four hours each week (two hours per session), which does not give these learners a chance to 'feel immersed' in English and acquire a habit of making form-meaning connections from the oral input they receive as much as the students who listen to the language for five hours a day (intensive group).

The fact that intensive exposure favors the development of foreign language listening skills can also be explained following retrieval theories from the cognitive psychology literature, according to which, if subsequent presentations of

an item are too spaced from the previous ones, no retrieval would be possible, since the item will no longer be in working memory and thus, no positive effect will be obtained from repetitions (Verkoeijen et al., 2005). As was previously mentioned, when language learners are constantly exposed to language patterns, the form-meaning associations of such patterns are strengthened (N. Ellis, 2002), yet if previous presentations of a specific pattern are not active when the pattern is repeated, no strengthening will occur. This situation is probably common in the case of the students in the extensive program, who are exposed to pattern repetitions within a minimum of a two-day lapse.

Apart from the cognitive explanation, there is another factor which can explain the greater improvement of listening skills in the intensive group. Since the sessions are especially long in the summer program, the teachers need to make a special effort to keep their students' attention; thus, more time is devoted to watching videos or listening to CD's than in regular courses. Additionally, the students in the intensive group are exposed to more output from their teachers, since more interaction occurs in these programs, again due to the length of the sessions (for example, during breaks, teachers tend to spend time with the students and have social interaction with them while having coffee). As a result, the students obtain more practice in listening comprehension than their peers in the extensive program and even in the semi-intensive. Apart from getting more practice, these students get 'better' practice, since it is more concentrated and more similar to naturalistic SLA or L1 acquisition, where automatization takes place easily due to the long periods of time the learners are exposed to the language

every day. Additionally, in intensive programs implicit learning is also promoted which undoubtedly must occur at the level of more frequent patterns which are strengthened through communicative interaction.

The positive results obtained by the students in intensive courses (both the semi and the intensive) in listening comprehension skills is in agreement with the teachers' beliefs about their students' progress in such groups. In the questionnaire distributed to the teachers, most of those who provided instruction in intensive courses claimed that the students following those programs acquired superior listening skills as compared to their peers in regular courses. Moreover, these teachers also admitted that more listening practice is provided in intensive programs. With respect to the students, when they were asked whether intensive courses were superior to extensive in terms of promoting listening skills, only half of them agreed with that statement. Furthermore, some students even believed more gains in this area could be obtained in extensive programs, which is certainly unexpected, given the prominence listening activities receive in intensive courses. The students in the semi-intensive group seemed to believe, for the most part, that they improved their listening skills more in that type of course than in extensive courses.

Concerning **grammar skills**, the measures within the written test that most closely examine this area are the sentence conversion exercise and the cloze task[49]. In the former activity the students are required to possess a good overall knowledge of the grammar of English in order to be able to express the same

[49] The cloze activity which was used in this study for level 3 did not target lexical items. Only grammatical knowledge was tested (see Appendix A).

meaning as in the sentence they are given, employing different forms (yet the learners only have a limited choice, since they are provided with the beginning of the sentence they have to create). The students' role is to produce a totally new sentence; in contrast, in the cloze activity they are only required to complete sentences with one word, task which can be said to demand fewer processing efforts than creating sentences.

The results of the analyses performed in this research study showed that, for the sentence conversion exercise, the same as for the listening activity, better and better results were obtained as intensity of exposure increased, with more progress observed in the students in the intensive program, followed by the students in the semi-intensive and then the extensive (see Figure 6.2 and Table 6.2). The results of the cloze activity, on the other hand, were not significantly different across program types (see Figure 6.3 and Table 6.2).

There have not been many studies investigating learners' improvement of grammar skills in intensive programs and the few studies which have investigated this issue have reported significant improvement in this area after completion of an intensive course (Schneider, 1977). Moreover, when comparing students in intensive and regular foreign language programs with respect to grammar skills, a more considerable progress has been reported in the case of the former (Frank, 1972).

As previously stated, the sentence conversion exercise requires a more creative and advanced use of the language than the cloze activity. The former exercise is more cognitively demanding than the latter; that is why conditions

which favor strengthening and retrieval of grammar structures should be more beneficial for this type of activity than for other types for which less cognitive effort is necessary. In intensive courses, due to the more concentrated exposure to grammatical patterns it is easier for those patterns to be strengthened and later retrieved than when exposure is so widely spaced that access to previous presentations of grammatical structures may be hindered or impossible. Nonetheless, when the task the participants need to perform is not so cognitively demanding, the advantage for concentrated exposure may not be so obvious. Appleton et al. (2005) demonstrate in their study that when subsequent presentations of an item are not so easily retrieved due to either increased spacing between presentations or changes in format, *memory* for such material suffered in *cued-recall* tasks; however, *recognition* of previously presented material in such difficult circumstances (which is less cognitively demanding) did not suffer as much. These results are comparable to the results in the sentence conversion exercise and the cloze. In the sentence conversion, the students in the intensive group had an advantage over the extensive group due to shorter spacing between presentations of grammatical patterns (both explicit and implicit), which facilitated the access and retrieval of previously presented material. This advantage is similar to the advantage experienced by the participants in Appleton et al. (2005) in cued-memory tasks. For the cloze test, which is less cognitively demanding (the same as the recognition task included in Appleton et al., 2005), the fact that the presentations of grammatical structures in the extensive program were too widely spaced did not hinder subsequent retrieval as much as in the more cognitively

demanding sentence conversion exercise. The fact that the sentence conversion is a more difficult task than the cloze is demonstrated by the students' mean scores in those tasks (Table 6.1), which are clearly higher in the later than in the former for the students in all the program types.

In general, it can be said that there is a certain advantage for students in intensive courses (both semi and intensive, but slightly more obvious for the latter) as compared to those in the extensive group in terms of grammar skills. Such advantage is not as clearly expected as for listening comprehension skills, since the students in intensive programs do not get more grammar practice than their peers, but quite the opposite. However, it can be the case that these learners are implicitly registering the grammar patterns to which they are exposed in the oral input. According to the teachers' opinion, in general, the students in intensive courses learn as much grammar as their peers in extensive groups (although for a few teachers they learn more). The students in the intensive program are also happy about their progress in grammar, as more than half of them claim to have learned more with more concentrated hours of instruction. The majority of the students in the semi-intensive group also report that they learned more grammar in such course than in traditional ones.

In terms of **vocabulary**, the results of the exercise reading B could be taken into account in order to explain the progress made by the students in the three types of programs under analysis. This activity required the students to look for synonyms of some words provided in the paragraph indicated[50] (see Appendix A,

[50] Apart from retrieval of previously encoded vocabulary items (which may or may not be

part 4). The scores obtained by the students in the intensive program in the reading B task were higher than those obtained by their peers in the extensive and the semi-intensive course, as shown in Figure 6.5 and Table 6.2. Such table indicates that there is a significant difference in the *program x time* variable, indicating that the program types under consideration evolved through time in a different way, with the learners in the intensive group starting with the lowest scores in the prestest and obtaining the highest in the posttest. That intensive foreign language courses help the acquisition of vocabulary has been reported by previous studies on intensive language learning under the ASTP (Leavitt, 1943; Oswald et al., 1950), as well as by reports on study abroad programs (Ife et al., 1998; Lennon, 1990, Milton & Meara, 1995) and intensive English in Canada (Collins et al., 1999).

The acquisition of vocabulary has also been analyzed by a number of studies in the cognitive psychology literature (Bahrick & Phelps, 1987; Pavlik & Anderson, 2005). This research has shown that spacing presentations is better for long-term memory, even if performance in immediate recall tests might be better for massed exposure. However, when comparing the 'massed conditions' in the experiments in the cognitive psychology literature with the more concentrated type of exposure in the programs under analysis, it can be said that 'massed' does not correspond to intensive language learning. Despite the fact that the hours of instruction in intensive courses are concentrated, no massed repetitions of items occur in the same way as in the cognitive psychology experiments, since there are

---

the case, depending on whether the students had been previously exposed to the word they were supposed to provide), another skill that is tested with this exercise is the ability to infer meaning from context.

always words intervening between target words, or grammatical structures intervening between target grammatical patterns. The language programs under consideration can all be said to provide distributed exposure to L2 forms. Nonetheless, if a parallelism is to be established with the cognitive psychology literature and the different learning conditions in the foreign language courses under analysis, it could be argued that the presentations in extensive courses are so widely spaced that such condition makes it hard for a subsequent presentation of an item to retrieve a previous instance. On the other hand, the less widely spaced presentations in the intensive program may facilitate retrieval, because when presentation 2 (P2) of an item appears, there is still a memory of presentation 1 (P1), and P2 will make P1 stronger. Consequently, the fact that vocabulary acquisition has been shown more advantageous for students in the intensive program is not in disagreement with the findings reported in the cognitive psychology literature.

The positive results in terms of vocabulary for the more intensive groups are consistent with the students' own beliefs about their vocabulary learning in such courses. While many teachers in intensive and semi-intensive groups did not generally believe that the students learn more vocabulary in such programs than in regular courses, the great majority of the students in intensive/semi-intensive groups feel their vocabulary knowledge improved more in these courses than in extensive courses.

With respect to **reading comprehension**, the results of the reading A exercise did not provide any advantage to the students in the intensive group, in

contrast with what was reported for other language areas in this section. Although there were no significant differences between the three program types in the posttest reading comprehension activity, the students in the semi-intensive group were the ones who demonstrated more progress in this exercise; followed by the students in the extensive group and then the intensive (see Figure 6.4 and Table 6.2).

Contrary to the results obtained in this study, there have been others which have shown superior reading skills for students in intensive language courses as opposed to regular courses (Frank, 1972; Lapkin et al., 1998; Lightbown & Spada, 1994; Peters, 2000). The fact that no similar results were obtained in the case of the learners in the intensive program under consideration could be attributed to the fact that less time is devoted to reading practice in intensive programs, as reported by many teachers in the questionnaire that they completed. Since most reading tends to be performed at home rather than during class time, it can be assumed that, in a longer period of time (such as seven months in the case of the extensive program) more reading can be practiced than in just one month (which was the time allocated for the intensive course). Interestingly, and contrarily to what was reported for other language skills, not as many students (only half) in the intensive and semi-intensive programs report having improved more their reading comprehension skills in those courses than in regular classes.

Summarizing, the investigation of the progress experienced by the learners in this study as analyzed by their results in a test which targeted listening, grammar, vocabulary and reading skills, seems to suggest that there is an effect of

time distribution of instruction hours on learners' language gains, as observed by the more skillful performance of the students in the intensive program when compared with their peers both in the semi-intensive and the extensive groups in most language areas except for reading comprehension. Such results could be explained by cognitive psychology theories concerning retrieval of previously encoded presentations, as well as by the amount and type of practice provided in each program for each of the language areas under research.

7.1.2. Time distribution and performance in writing

Concentrated time distribution can be said to have a certain positive effect on students' writing skills, in view of the results presented in section 6.1.2. Considering the programs under analysis, when the hours of English instruction were distributed over a longer time period (**extensive program**), students' writing skills significantly improved from the beginning until the end of the course in terms of fluency (as measured by words per T-unit), and syntactic complexity (as measured by clauses per T-unit). However, some trade-off effects were also found for this group, since the learners' performance with respect to lexical richness and accuracy was significantly poorer in the posttest than in the pretest.

Similar trade-offs have been reported in several studies dealing with second language learning (not necessarily in intensive courses) in written and in oral production between fluency and accuracy (Yuan & Ellis, 2003), or complexity and accuracy (Mehnert, 1998; Skehan & Foster, 1997), especially at beginning stages,

due to the students' lower processing capacity, which makes them focus on some aspect/s of the message while not devoting so much attention to others. The trade-offs observed for the students in the extensive program relate to accuracy in favor of more fluent and grammatically complex production which is, at the same time, less lexically rich. Improvement in other language areas in detriment of accuracy has been widely reported (Mehnert, 1998; Skehan & Foster, 1997; Yuan & Ellis, 2003).

On the other hand, the lack of progress in lexical complexity can be due to the type of task that the students were asked to complete. The topics proposed in the pretest ('My best friend') and in the posttest ('Someone I admire') are highly similar, which helps comparison between pre and posttest; nevertheless, this design could limit the students' use of a wide variety of vocabulary items, due to the fact that many similar words are expected to be produced in both compositions; in fact, many students wrote about their best friend when describing somebody they admired.

Regarding the students in the **semi-intensive group**, no significant progress was observed from pre to posttest in any of the measures used to analyze written production. Indeed, the students' compositions were significantly less lexically complex at the end of their course than at the beginning, which could be explained, again, by the type of task that they were required to perform, as well as by the low emphasis the teaching and practice of vocabulary usually gets in the classroom (Folse, 2004). Since no significant progress was observed in any of the measures, there were no trade-offs. Accuracy does not become significantly poorer in the

posttest, as happened for the students in the extensive program; however, contrarily to these learners, the students in the semi-intensive group did not make significant progress in fluency or syntactic complexity.

The behavior of the students in the **intensive group** is slightly different from what has been reported for the learners in both, extensive and semi-intensive programs, which is why concentrating the hours of instruction can be said to have a certain positive effect on the acquisition of L2 writing skills. First of all, the students in the intensive group were significantly more fluent in the posttest than in the pretest, while also using more syntactically complex language. Additionally, these learners used significantly more complex vocabulary, as opposed to what occurred in the case of the students in extensive and semi-intensive groups, whose compositions were significantly less lexically rich in the posttest than in the pretest. Consequently, even if the task the students were requested to complete in order for their writing skills to be assessed encouraged, to a certain degree, the use of highly similar (and highly familiar) vocabulary in the pretest and in the posttest, the students in the intensive group were more creative than their peers and used more complex lexical items at the end of their course than at the beginning. The more advanced performance in terms of vocabulary demonstrated by the students in the intensive program can be explained in similar terms as for the vocabulary activity included in the written test (see section 7.1.1) Finally, although accuracy did not improve from pre to posttest in the case of the students in the intensive group, it did not become significantly poorer, as was found for their peers registered in the extensive group. As a consequence, it can be said that the students in the intensive

program demonstrated more progress in their writing skills, since more areas improved from pre to posttest than in any of the other groups (fluency, syntactic complexity and lexical richness) without such improvement having a highly negative effect on these students' accuracy.

The analysis of the acquisition of writing skills in intensive programs has not been popular due to the fact that, since the beginning of the implementation of such programs, speaking and listening skills have been given more prominence. In the intensive courses under the ASTP, writing was a secondary skill and was not given much attention because the focus of such programs was the development of audio-oral skills. In the analysis of other intensive programs which were not related to the ASTP (which did not relegate writing as much) not many tests that measure students' writing skills have been used. Nevertheless, there have been a few authors who have reported superior writing skills in the case of students registered in intensive programs. For instance, McKee (1983) observed that the learners in her study that attended an intensive summer French program showed better listening, reading and writing skills than those in the regular French course; however, the only area where the students in the intensive group showed a *significantly* more skillful performance was in creative writing. McKee (1983: 20) concludes that the students in the intensive program are "more eager to use the target language (at least in written form) than students in the traditional classroom." Similarly, Lapkin et al. (1998) in their study of three different models of French instruction (the half-day model (half-day of L2 instruction in 10 weeks), the 80-minute model (80 minutes a day for 5 months) and the regular, which offers

40 minutes a day for 10 months) observed that only the learners in more concentrated models (half-day and 80-minute) made significant gains in writing skills (as demonstrated by the scores obtained in two compositions) from the pretest to the posttest, with more gains being made in the most concentrated model. Germain, Netten, & Séguin (2004) also report significantly higher performance in writing for the students in intensive French as opposed to those in regular French classes.

The fact that the students in the intensive program under analysis demonstrated more improvement in writing skills was not so clearly expected, taking into account the amount of practice devoted to this skill in class, which is higher in the extensive, then in the semi-intensive and less commonly practiced in the intensive, as reported by the teachers and the students. Typically, the students write a composition every week or every two weeks in the extensive course, while two are assigned every week in the intensive course, which adds to a highly different amount when considering the weeks of class in the extensive and the intensive programs (a total of up to 20-25 compositions can be expected in the former and 8-10 in the latter). Nevertheless, the concentrated input to which the students in the intensive program are exposed and the more continuous practice of oral production skills (as will be explained in the next section) probably also helps writing skills. It must also be remarked that questions of style or other areas which are more clearly related to quality of writing (structure, cohesion, paragraph development, etc.) were not analyzed, and those are probably the skills for which more practice should have had an effect in the case of the students in the extensive

program; however, such differences could only be perceived through more qualitative analyses. Since the measures used in this study were quantitative and referred to gains in fluency, syntactic and lexical complexity and accuracy, those areas could have been fostered by practice in listening, or, especially speaking skills in the case of the students in the intensive course.

Taking into account the results obtained by the students in the measures used to analyze gains in writing skills, one of the findings which might have been unexpected is the fact that the students in the semi-intensive group do not demonstrate any of the advantages claimed for learners in intensive programs in general or for the intensive program under study in particular; moreover, the semi-intensive students' gains in certain writing measures were lower than for the students in the extensive program. This result can be explained considering that the students in the semi-intensive program have more hours of instruction a week than those in the extensive model, yet the sessions have similar duration (2 hours for both the extensive and the second semester semi-intensive, and 2.5 hours in the first semester semi-intensive). Maybe the determining factor in terms of intensity in language learning is not the amount of hours a week, but per session, yet more research should be performed in order to verify such suggestion. Additionally, the students in the semi-intensive group have to dedicate their time not only to their English class (which they do for 8-10 hours a week, plus the hours of homework), but also to the rest of the university subjects which they are taking at the same time, as opposed to the students in the extensive group, who only dedicate 4 hours to English, or the students in the intensive group, who usually only have English to

worry about, in terms of academic subjects. Indeed, it has been highlighted by several authors that one of the reasons why intensive courses may appear to foster higher levels of proficiency is the fact that the students can dedicate a block of time just to the learning of the foreign language and not be distracted by other subjects in their degrees (Kalivoda, 1975; Kilker & Gunderson, 1972; McKee, 1983; Schulz, 1979). It appears that the students in the semi-intensive group may be at a disadvantage with respect to both groups; with the students in the extensive group, because not only do they have to cope with their academic subjects, but additionally they have English class every day; and also with the students in the intensive group, because they do not have the privilege to concentrate exclusively on English, and they do not benefit from a high amount of hours of instruction per session.

7.1.3. Time distribution and performance in speaking

The students' progress in speaking skills was observed to be slightly different according to the way the hours of instruction in the program where they were registered were distributed, as reported in section 6.1.3. Although no significant differences existed in the posttest between the students in the extensive, semi-intensive and intensive programs, the progress that the learners experienced appeared to be different for each type of language course.

The students in the **extensive** group made significant progress with respect to their fluency and their syntactic complexity, as measured by the syllables per

minute and clauses per T-unit respectively; on the other hand, lower scores were obtained with respect to accuracy in the posttest than in the pretest, despite such difference not being significant. As was noticed in the case of writing, trade-off effects are also present in oral skills for the students in the extensive group, who, while significantly improving their fluency and syntactic complexity, showed less accurate oral production in the posttest. The students' performance in terms of accuracy can be explained by the fact that these learners were trying to produce longer and more complex T-units, which were in consequence more erroneous because they were less automatized. The poorer production in terms of vocabulary can be explained in the same way. An additional reason why no improvement is observed in vocabulary is the fact that the same task was used in the pretest and in the posttest; as a consequence, not much chance was given to the students to supply a wider range of lexical items at the end of their language course.

The students in the **semi-intensive** program only made significant progress from pre to posttest in terms of syntactic complexity. Again, one area in oral production improves (syntactic complexity) at the expense of others (fluency, lexical richness and accuracy), due to the students' limited processing resources and their allocation to one area in detriment of others.

Finally, the students in the **intensive** program demonstrated significant progress from pre to posttest in all the areas analyzed for oral production (fluency, lexical complexity, and accuracy) except for one (syntactic complexity). With regard to this group of students, some trade-offs are also observed due to the fact that automatization is still not complete for these learners either (even if the

students in the intensive program show more skillful oral production than the learners in the other groups), and cognitive resources cannot be allocated to all the aspects of oral production. There is only one area where improvement is not made by the students in the intensive group, as opposed to three for which a significant progress was observed (in contrast with the students in the extensive program for which progress in two areas had detrimental effects on two, or the students in the semi-intensive course, for whom progress in one area had detrimental effects on two and not much impact on one). The fact that the students in the intensive program did not show significant progress in syntactic complexity can be explained through the allocation of cognitive resources to other areas of oral production, as suggested above, or by the fact that the task in the pretest was the same as in the pretest; therefore, the lack of cognitive complexity may account for the lack of increased syntactic complexity (Robinson, 1995; 2003b).

That the students in the intensive program made more progress than their peers in less intensive courses in oral production does not come as a surprise, considering previous research on the topic. Since audio-oral skills were the focus of the intensive courses under the ASTP, highly positive results in terms of speaking skills were obtained by the students following this program in the Army and similar models in colleges (Agard, 1946; Agard et al., 1945; Goedsche, 1946; Leavitt, 1943), such positive results being more evident with respect to fluency, since accuracy was not emphasized (Pargment, 1945). Similar achievements in speaking skills, especially in oral fluency, have been widely reported for the students in French immersion programs in Canada, despite the fact that the lack of accuracy in

grammar and morphology has also been stressed (Genesee, 1987; Swain & Lapkin, 1986) or the less evident improvement in terms of complexity (Billy, 1980). Advantages in oral production skills are also reported for the students in intensive French in Canada as opposed to their peers in regular French as a foreign language classes (Germain, Netten, & Movassant, 2004). The literature on study abroad has also reported that the area in which the students improve the most is oral fluency (DeKeyser, 1991; Freed, 1995; Lafford, 2004; Segalowitz & Freed, 2004). Likewise, studies which have investigated the development of oral skills in intensive English courses in Canada have also found that concentrated L2 instruction favors the improvement of students' oral skills (Collins et al., 1999; Lightbown & Spada, 1994; Spada & Lightbown, 1989; White & Turner, 2005).

As was mentioned before, in terms of speaking skills, fluency was the area in which the students in intensive courses were reported to have improved the most. However, other studies have also emphasized the fact that students in intensive courses, apart from being more fluent than their peers following traditional foreign language instruction, are also more accurate (Lightbown & Spada, 1994; Spada & Lightbown, 1989). In the case under research, students in both, the extensive and the intensive program make significant gains in fluency, yet it is only the latter group which makes significant gains in accuracy.

The data obtained from teachers and students through questionnaires revealed that, in fact, both groups tend to believe that more progress is made in speaking skills in intensive courses than in extensive courses. Additionally, some teachers admitted that more speaking practice was provided in intensive courses

than in extensive courses; nevertheless, the responses were highly similar for students in intensive and semi-intensive programs and not so highly positive effects were obtained for oral production skills in the semi-intensive program.

The role of practice, thus, needs to be considered when explaining the greater progress experienced by the students in the intensive group, first in terms of 'quantity' (probably more hours were devoted to the practice of oral skills in the intensive group), but also regarding 'quality'. First of all, oral practice is probably provided in a wider variety of contexts in intensive courses than in extensive. The more time the students spend with their teachers per session, the more interaction takes place, not only in the class, but also in other contexts, such as taking a coffee during breaks, which is typical of intensive programs. The 'quality' of practice is also different in terms of the different time distribution of such practice in the program types under analysis. Through more concentrated exposure, the students have more chances to proceduralize the knowledge they acquire in their classes, in a way which is similar to naturalistic SLA or L1 acquisition. In naturalistic SLA, as well as in L1 acquisition, proceduralization is facilitated by the constant and massive amount of hours that the learners can devote to practicing their oral skills. The declarative knowledge which the students in the intensive program are acquiring is readily available and easily retrieved for oral practice, which is not so clearly the case for the students in the extensive groups, for whom the declarative knowledge required for oral practice may have already been forgotten due to the wide spacing between sessions. More automatization of oral skills has occurred for the students in the intensive group, as reflected by the fewer trade-offs experienced

by this group. Since these learners' oral skills are a bit more advanced, they can concentrate on several aspects of the oral task (fluency, lexical richness and accuracy) at the same time.

As was the case for writing, the advantages of concentrating the hours of instruction are not experienced by the students in the semi-intensive group. Again, the explanation may be related to the fact that intensity is significantly lower than in the summer intensive course, with sessions being similar in length to the extensive program. Additionally, the students in the semi-intensive course have to deal with their coursework, and are exposed to a lot of complex declarative knowledge which they have to acquire, the learning of which may interfere with the declarative knowledge they acquire in their English class. This might be one reason for the apparent failure in the proceduralization of oral skills by this group of learners, namely that the declarative knowledge is not so readily available for these students when they do oral practice.

7.1.4. Summary: Time distribution and its effect at the intermediate proficiency level

The results of this research have provided some evidence for what had already been suggested by many studies, to be precise, that when students' knowledge of the foreign language is at an intermediate stage, concentrating the hours of instruction can have positive effects on the acquisition of different language skills. It is true that the higher scores obtained by the students in the

intensive program under analysis may not be as significantly higher with respect to students in non-intensive language courses as some studies have shown (Collins et al., 1999; Lightbown & Spada, 1994); nevertheless, the intensive courses that these authors analyzed not only concentrated the hours of instruction, but they also offered more hours than the regular L2 courses. The results obtained here may be more 'modest' and more comparable to those reported by Lapkin et al. (1998), which referred to programs with the same hours of instruction and revealed a less strong superiority of the most concentrated model. Moreover, although the courses under analysis included 110 hours, there were only about 80 hours between the pretest and the posttest; consequently, that some progress was experienced by the students in such short period of time is highly significant, and more so that there were differences in favor of one program type (intensive).

The students' impressions on their gains usually resembled their actual progress (see Appendix C for the background questionnaire, especially questions 11, 15 and 16). In general, the great majority of the students in the intensive program preferred it to the more traditional EFL class and they mostly agreed that they learned more in an intensive course. Moreover, when completing the course, the majority of the students claimed that they had learned 'a lot' and there were even some who said they had learned 'really a lot'. However, the question in which agreement was the highest referred to the level of difficulty of the intensive course, for which an overwhelming majority of the students responded that intensive programs were harder than regular programs. The level of difficulty as well as the fact that intensive instruction was considered more 'enjoyable' may

have made the students work harder in their intensive class and therefore obtain significantly better results.

Another idea which has been suggested is that the positive effect of concentrating the hours of instruction may not be linear, according to how many hours a week the students attend their L2 class. It was observed that the students in the semi-intensive program not only demonstrate fewer language gains than those in the intensive program, but they also show some disadvantages, regarding written and oral production, with respect to their peers in the extensive program. The fact that the students in the semi-intensive program did not make as many gains as those in the intensive was explained in terms of the lower intensity per session, as well as by the fact that the students in the semi-intensive group could not concentrate on their English skills as much as their peers in the intensive program, since they had to work on their other university courses. Although there are many students in the semi-intensive program who thought they had learned 'a lot', there are more who thought that they had learned 'an ok amount' (as opposed to what was reported for the students in the intensive course, who mostly believed they had learned 'a lot'). Nevertheless, the students in the semi-intensive group still prefer such course to the extensive, and think that they learned more in the semi-intensive group (which does not correspond to the students' actual results). In terms of the difficulty of the program, not as many students in the semi-intensive course as in the intensive considered semi-intensive instruction harder than regular English lessons (half of the students thought the program was harder). Moreover, there were a few students (about a quarter) who believed the

level of difficulty in semi-intensive courses was lower than in extensive programs. In general, the students' attitudes and beliefs with respect to the semi-intensive program do not correspond to their results, since not much progress was observed in this group even if the students liked the program and considered they learned more than with extensive lessons.

Regarding the students in the extensive group, they generally believed they had learned less in their course than what the students in semi-intensive and intensive programs reported; nevertheless, more gains were observed in this group than in the semi-intensive in writing and speaking skills (but not in listening, grammar, reading and vocabulary). The reason for this outcome is probably due to the disadvantageous 'external' circumstances for the students in the semi-intensive group (as was mentioned before) rather than to an advantage of a widely spaced time distribution of hours of instruction in the extensive group.

7.2. Level 5 (advanced)

7.2.1. Time distribution and performance in listening, grammar and vocabulary exercises

The results of the test at the advanced level revealed that time distribution did not have any effect on the students' acquisition of listening, grammar skills, or vocabulary, as measured by a listening exercise, a sentence conversion, and a cloze activity. No significant differences were observed between the two program types

under consideration (extensive and intensive) in the pretest or in the posttest for any of the tasks. Furthermore, the students in the two programs made significant gains at the end of their respective course in listening, grammar and vocabulary, as reflected in the results of the listening exercise, and the cloze activity[51].

It has been previously mentioned that there have not been many studies analyzing the students' progress in intensive courses as opposed to regular courses at the advanced level. When evaluating the intensive programs under the ASTP, Paulsen (1945) suggested that intensive courses should be especially good at the advanced level, more than at beginning stages; however, he does not provide any empirical data or a definition of 'advanced'. On the other hand, Schueler (1944) and Springer (1944) claim that beginner or intermediate-level students benefit more from an intensive course than those learners whose knowledge of the L2 is already advanced. To my knowledge, there has not been any study comparing the performance of advanced students in intensive and regular foreign language courses in terms of listening, grammar and vocabulary skills. Nonetheless, studies analyzing students' gains in different types of intensive language programs depending on the students' initial proficiency level have often shown that the learners with lower proficiency tend to make more progress than those who are more skillful in the L2 at the beginning of the program. Such results have been reported in the study abroad context for general language proficiency (Freed, 1990), vocabulary (Milton & Meara, 1995), or listening skills (Lapkin et al., 1995). One reason which has been presented for the higher improvement in the case of

[51] The cloze included in the level 5 test did not only target verb forms and other functional words, as was the case of the level 3 cloze, but also some lexical words (see Appendix B).

students at lower proficiency levels as compared to advanced learners refers to the type of tests used to collect the data, which may have provoked ceiling effects for students at more advanced levels of proficiency (Freed, 1998).

That fewer gains are made when students are at a more advanced level is easily explained by the fact that there is less to gain when one already has a lot. When considering initial skill, gains are proportionally smaller at the more advanced initial levels; that fact can explain why progress is less apparent when the initial skill level is already advanced (Gardner et al., 1977). If a student learns 50 words when s/he already knew 1,000 less gain will be evident than if s/he just knew 100 words to start with.

Nevertheless, the scores on the test used for this particular research show significant gains in listening, grammar and vocabulary skills at the advanced level for both program types, as reflected in the learners' performance in a listening and a cloze exercise. What the studies analyzing gains in intensive courses according to initial proficiency level may illustrate, however, is that intensity may not be so clearly beneficial when the students' performance in the L2 is already advanced.

A possible interpretation for these findings could be that the type of improvement in the language skills under consideration that students at more advanced proficiency levels are supposed to make is not affected by the type of exposure which was provided in the intensive summer program under consideration (110 hours in one month). Intensive exposure in a short language course may be beneficial at lower proficiency levels due to the greater possibility such exposure offers to strengthen and practice frequent grammatical patterns,

collocations, or lexical items, by providing fresh declarative knowledge that the learner can better proceduralize. Nevertheless, in the case of the acquisition of more advanced linguistic material such exposure should have no effect, since proceduralization has already taken place at that level (frequent patterns) and further acquisition of the different aspects of the language is usually a matter of acquiring new declarative knowledge about different aspects of a grammar rule the students already know and use, or about specific vocabulary items or collocations, which are not so frequent in the every-day input.

The results observed for the listening, grammar and vocabulary skills at level 3, which favored semi-intensive and intensive groups (especially intensive) seemed to suggest that at that level of proficiency, when the students still need to proceduralize the knowledge about new grammar rules, vocabulary items, etc., having such knowledge fresh in their mind (which occurs in intensive classes) helps proceduralization[52]. The case of the students at the advanced level (who have already proceduralized much of their L2 knowledge and only need to learn highly specific aspects of the L2 through declarative knowledge, or repair wrong forms which they have already automatized) relates to the acquisition of new declarative

[52] It is debatable whether the knowledge the students evidenced when completing the tasks chosen for this study is declarative or procedural. As R. Ellis (2004) suggests L2 learners typically use both explicit (or declarative) and implicit (or procedural) knowledge when completing a task. Whatever the claim might be, what appears less controversial is the fact that in order to perform the activities included in this particular study the learners needed access to the explicit and implicit knowledge they had been acquiring throughout their language course, and the students in the intensive (and also semi-intensive) intermediate classes seemed to have such knowledge more available in terms of listening, grammar and vocabulary than those in extensive groups.

knowledge, the proceduralization of which will take more time than for the declarative knowledge at lower proficiency levels. Proceduralization of advanced and less frequent language patterns (which are not often present in the every-day input in the communicative situations that take place in intensive courses but not as often in extensive programs) could possibly benefit from intensive language exposure, but not from an intensive course which offers as few hours of instruction as the one under study (110 hours), which, nonetheless, clearly benefits proceduralization at a lower proficiency level. Students at the advanced level could benefit from exposure to the target language in a naturalistic context, where they would have a chance to continue with the proceduralization process, which would lead, with practice, to automatization (DeKeyser, 2007b). When faced with natural input learners have more possibilities of encountering less frequent structures or vocabulary than in the classroom; moreover, the level of intensity is higher. Alternatively, advanced EFL students could possibly obtain more advantages from a longer intensive course which targets more specifically the areas in which improvement is still necessary for advanced learners.

When considering the students' opinion about the intensive program (see background questionnaire in Appendix C, questions 11, 15 and 16), the majority of the students claimed to have learned more grammar in other previous non-intensive courses than those who said they learned the same or more. For vocabulary, the opposite is found, with an overwhelming majority of the students reporting that they had learned more vocabulary through concentrated instruction. The teachers' opinion about their students' progress in the different skills is more

or less in line with the students' opinion in terms of grammar, since they did not suggest any advantage in this area for the students in the intensive program; nevertheless, regarding vocabulary, the instructors tend to think that students in the intensive program learn as much vocabulary as their peers in non-intensive groups. In general, it can be said the students' opinion about their progress in the intensive program with respect to listening and vocabulary tends to be a bit more optimistic than what is shown in their test results. As for grammar, the students in the summer intensive program did not consider that program type superior to the extensive in terms of fostering grammar skills. This opinion is more in agreement with the results obtained by these learners in grammar, which were highly similar to the results in the extensive group.

7.2.2. Time distribution and performance in writing

The effects of time distribution on writing at the advanced level do not seem to be as obvious as in the case of intermediate learners, considering the results of the statistical analyses performed for the level 5 learners in the extensive and the intensive EFL programs in the measures which were chosen to analyze gains in fluency, syntactic and lexical complexity, and accuracy.

There is certainly a dearth of studies comparing the acquisition of writing skills at the advanced level by students in intensive and regular courses. Keilstrup (1972) claims greater gains in writing and oral skills in an intensive German course for advanced students than for those attending regular German lessons.

Nevertheless, no actual tests or results are provided, and more importantly, what the author considers 'advanced German' would be the equivalent of four semesters of German instruction in college, which would be similar to the level 3 (intermediate) in the programs analyzed in this study.

According to the results presented in section 6.2.2, it could be argued that time distribution had a certain positive effect on the students' development of writing skills in the extensive program, since the learners in such program significantly improved their performance in terms of fluency (words per T-unit) and syntactic complexity (clauses per T-unit), while the students in the intensive course did not demonstrate significant gains in those areas, or any other (see Table 6.10). Conversely, the results presented in section 6.5.1, which compared the performance of the advanced learners with native-English speakers (NESs), revealed that, while the scores on the pretest for the students in the intensive program were only significantly different from the scores of the NESs concerning accuracy, the students in the extensive group differed significantly from both the intensive learners and the NESs regarding written fluency. Moreover, despite the fact that the students in the extensive and intensive programs were not significantly different in terms of syntactic complexity in the pretest, the learners in the intensive program had a higher mean of clauses per T-unit (2.19) than their peers in the extensive program (1.98), but also than the NESs (2.09). Consequently, it cannot be assumed that time distribution accounts for the fact that the learners in the extensive program demonstrated greater progress in fluency and syntactic complexity than those in the intensive course. Such progress was rather due to the

fact that the students in the extensive group had some room for improvement until achieving native-like performance in those measures, while their peers in the intensive group did not.

The results for the other measures under consideration (lexical complexity and accuracy) were highly similar for the students in the extensive and intensive programs in the pretest and in the posttest. In terms of vocabulary, the performance of the advanced students is not so distant from that of the NESs in the pretest (section 6.5.1). Another finding reported in section 6.2.2 reveals that no improvement was observed in lexical complexity by any of the groups; moreover, the posttest was less lexically rich (though not significantly) for all the students. Such result can be a reflection of the nature of the task the students had to perform: writing about their best friend in the pretest and someone they admired in the posttest. Not only are these two tasks highly similar in terms of the vocabulary that the students are supposed to use, but also they are not cognitively challenging for students at this level of proficiency (more complex tasks have been shown to lead to more complex performance (Robinson, 2003b)). Whereas having the same task for intermediate and advanced students is helpful in order to perform comparisons between the two groups, such arrangement is not beneficial if the students are to be challenged at more advanced proficiency levels.

With respect to accuracy, the results of the accuracy measures (EFTU/T and Err/T) demonstrate that both groups of learners (in extensive and intensive programs) need to improve in that area a great deal before reaching native-like performance (section 6.5.1). Moreover, the results reported in section 6.2.2 indicate

that the performance of the learners in the two level 5 programs was not more accurate in the posttest than in the pretest. The fact that a significant improvement is not observed in terms of accuracy is probably related to the automatization of some non-target forms which has occurred for some students at the advanced level, many of which reflect L1 transfer (confusion between *do*/*make*, use of *-ing*/*to infinitive,* inverted word order in embedded *wh*-clauses as in *I know what is it like*, lack of third person singular *–s* in simple present, etc.) In order for the students to eliminate these forms, instruction must concentrate on those grammar/vocabulary points; however, the EFL advanced courses under study tend to emphasize other areas of language learning (speaking and listening skills, especially in the case of the students in the intensive program) in detriment of other areas, such as grammar, for which a highly skillful knowledge on the part of the students is already assumed, at least in the areas in which some students still seem to be making errors.

The data from the questionnaire that was distributed among students and teachers revealed that there is notably less writing practice in intensive courses than in extensive. In the extensive level 5 course the students are usually assigned a composition every week or every two weeks, thus, considering the number of weeks the course lasts, there can be a total of approximately 20-25 compositions. For the students in the intensive program to write such number of compositions, they would have to write one each day, which is not feasible for either students or teachers. As a consequence, apart from ceiling effects, another reason why the learners in the intensive program did not experience more progress as compared

with the extensive in writing may be that the former had much less practice.

Regarding the students' impressions about their course, half of the learners in the intensive program thought they improved their writing skills more in such program than in regular EFL classes, which is not in agreement with the actual results. The instructors' opinion, on the other hand, was closer to the concrete results; namely, they thought that the students in the intensive program make less progress in writing skills than those in the extensive course. In this case, as for listening and vocabulary, the students in the intensive group believe they learned more in such program than what their results in the different tests suggest.

7.2.3. Time distribution and performance in speaking

The results presented in section 6.2.3 suggest that time distribution had no effect on the acquisition of oral skills at the advanced level in the two programs under consideration. As was mentioned before, the effect of time distribution on the acquisition of speaking skills at the advanced level has not received much attention in the SLA literature. The studies that have included advanced students in intensive language learning contexts have only compared their progress with intermediate students in the same context, and not with advanced students in a regular program. The authors comparing advanced and intermediate learners in terms of progress in oral skills in intensive contexts have reported greater gains for those who started with a lower proficiency level in the case of oral fluency (Freed, 1995; Gardner et al., 1977; Lapkin et al., 1995).

In this particular research study there were no significant differences between the two programs (extensive and intensive) in any of the measures used to analyze oral production, suggesting that the two programs were comparable; furthermore, significant gains were made in only one area by the two groups, and such area happened to be the same: oral fluency (see section 6.2.3). The only difference that was remarked between the intensive and the extensive group (which was not significant) referred the trends observed for the two groups regarding syntactic complexity and lexical richness. While the students in the extensive program showed slightly worse performance in those aspects in the posttest, the students in the intensive group made some gains, although they were not significant. As a consequence, trade-off effects between fluency on the one hand, and syntactic and lexical complexity on the other, may seem more evident in the case of the students in the extensive program.

As was the case for written production, the fact that progress was not experienced by the advanced learners in some measures could be due to ceiling effects (6.5.2). Indeed the students' performance in terms of syntactic and lexical complexity in the pretest was comparable to the NESs'. However, while the scores in fluency and accuracy were significantly different between native and non-native English speakers, progress was only registered in terms fluency (not in accuracy) for the students in the two programs. Accuracy seems to be the most problematic language area for all learners at all proficiency levels. The same as in written production, inaccurate L2 forms in the oral production of advanced students are mostly automatized non-target forms (*childs, childrens* or even *sons* instead of

children, double negatives *they haven't nothing to eat,* problems with prepositions *is looking the basket, there is a dog on the house*). Additionally, there are other mistakes which could have possibly been avoided in the written mode but were made due to the fact that learners can concentrate less on forms while speaking than while writing because of the immediacy of the oral production (*they goes to the mountain*, *their mother are looking after them,* or *the dog is so happy because he has eat the food*).

As was claimed in the case of written production, the classes at the advanced level do not focus so much on the language forms that are supposed to be mastered at this level. The advanced program under study, according to the teachers and students involved, emphasized some skills (such as speaking or listening, especially in the intensive program) to the disadvantage of others, such as grammar instruction, especially on those aspects that the students are already supposed to master. Even if feedback is provided for non-target forms during speaking activities, neither the students nor the teachers tend to consider grammatical accuracy as the main point in oral production. The mistakes in speaking due to the lack of time the students have to concentrate on both form and meaning could probably be avoided through more oral practice.

Considering the fact that speaking skills have traditionally been targeted in intensive courses, and more practice is devoted to these skills in such courses (as was the case for this particular intensive group, as reported by the teachers and students themselves), more gains would have been expected for the students in the intensive group under analysis than for those in the extensive program. This lack of advantage in syntactic and lexical complexity is due to the fact that the learners'

performance in the pretest in both groups was already native-like; consequently, there was not much room for improvement or for differences between the two program types in the posttest. However, there is no difference between the students in the extensive and the intensive program in terms of the other two areas where improvement could be made, namely fluency and accuracy: both groups make significant progress in terms of fluency and no progress in accuracy. It could be the case that a more challenging task might have provided the students in the intensive group more chance to demonstrate their gains than the task used for this research, which was quite simple and even the learners at the intermediate level showed significant progress from pre to posttest. Possibly, a more complex task would have discriminated better between fluency at the advanced level. Similarly, having used different tasks in the pretest and the posttest might have increased task complexity and thus could have provided the opportunity for more fluent and complex performance in those more advanced learners. The case of accuracy has already been discussed in previous paragraphs.

The fact that the intensive program does not promote higher speaking skills than the extensive is not in agreement with the majority of the students' beliefs regarding such program. The students in the intensive course reported to have progressed more in their oral skills in the intensive course than in extensive courses. Also, the teachers claimed that these students perform better in this language area than their peers in the extensive group. It could be the case that, as suggested above, it was because of the task used to elicit oral production that the students in the intensive group were not able to show the full extent of their skills

in this area.

7.2.4. Summary: Time distribution and its effect at the advanced proficiency level

The results of the analyses performed for this study suggest that concentrating 110 hours of English language instruction in one month does not have any clear positive effect at the advanced proficiency level in terms of listening, grammar, vocabulary, writing or oral skills (contrarily to what was observed at the intermediate proficiency level). It has been proposed that the different stages in the acquisition process where the intermediate and advanced learners are, together with the difference in the type of material to be acquired, can account for the fact that, whereas intermediate students benefit more from the intensive language course under examination in most language skills under analysis, advanced students do seem to make comparable L2 gains in the intensive program. In terms of Anderson's skill acquisition theory (Anderson, 1993; DeKeyser, 2007b) it could be argued that the advanced learners under consideration have already been through the process of proceduralization of most skills in the L2. These learners have achieved a certain degree of fluency in their L2 use, in terms of being able to perform form-meaning mappings at a reasonable rate both in production and comprehension, which was attained by massive practice during previous years. What was expected from them in the level 5 course they followed was to increase their declarative knowledge of the grammar rules they

have already proceduralized by learning more specific applications of such rules, making new generalizations or analyzing exceptions. It could simply be that the proceduralization of such specific knowledge needs more practice than what can be obtained in a 110-hour course, whether it is intensive or extensive. Additionally, these advanced learners are expected to modify the non-target forms that they have automatized, which, again, would require more focused practice of those forms in different contexts, as well as explicit feedback on inaccurate production. It was suggested that, whereas a 110-hour course at the intermediate level could promote language gains and advantages for those students in an intensive course, at the advanced level (when qualitatively more complex patterns are expected to be learned) an intensive course which offers the same amount of hours of instruction does not seem to trigger the same positive effect in terms of students' gains.

It has also been suggested that, in the case of written and oral production, the advanced learners could not make much progress in some measures because their performance was already quite native-like in the pretest. Moreover, it was claimed that, whereas the advanced learners improved in the areas where there was room for improvement in the pretest, the case of accuracy was an exception: the students' performance is quite far from native-like and not much progress is experienced in this direction by the advanced learners in any of the two programs.

It has also been suggested that the students in the intensive course tend to be more optimistic about their progress in most language skills than what their actual results show. More than half of the students claimed to have learned more

in such course than in regular EFL courses, and they also liked intensive instruction more (yet, the number of students having this opinion at the advanced level in the intensive program is lower than for the intermediate intensive group). When asked more specifically about how much they had learned in the course they had followed, the majority of the students in the two program types under consideration answered 'ok' (2 in a 4-point scale); nevertheless, the next big group of students in the extensive group claimed 'little', whereas the second larger group in the intensive course reported to have learned 'a lot'. Additionally, the students in the intensive program overwhelmingly considered that program harder than regular English instruction.

Summarizing, despite the fact that the students in the intensive program mostly believed they improved their English language skills more in such program than they had in other extensive courses, the results of the analyses performed for this particular study did not confirm this assumption. There is a very slight advantage in two speaking measures for the students in the intensive program (syntactic and lexical complexity); however, the progress experienced in the other skills is highly similar when comparing the two program types. It was suggested that the lack of a clear positive effect in the case of the students in the intensive program at the advanced level was mainly due to the kind of knowledge the students have to acquire and the few hours of instruction offered. It was also suggested that some of the tests/measures may not have been adequate for discriminating between the gains made by the students in the two program types at the advanced level.

7.3. Discussion on the measures of written and oral production

The results of the comparison between intermediate and advanced learners in terms of the measures adopted in this dissertation to analyze progress in writing skills (section 6.4.1) and speaking skills (section 6.4.2) indicate that the performance of the advanced learners in all the measures (except for accuracy in writing) was significantly superior to the performance of the intermediate learners. These results indicate that the measures selected in this study can capture well the improvement experienced at the intermediate level. Moreover, the differences found between the three programs at level 3 can also be considered reliable, since the measures adopted in this study have demonstrated to discriminate between learners at different stages in their acquisition process. As was mentioned before, the students at the advanced level were not significantly more accurate in writing than their peers at the intermediate level in the pretest or in the posttest. This finding may indicate either that the measures selected to examine accuracy are not adequate to capture the progress experienced by the students from an intermediate to an advanced proficiency level, or rather that more errors are made by the advanced students because their T-units are longer and more complex (see section 7.2.2 for a deeper discussion on the issue of accuracy at the advanced level).

Considering the appropriateness of the measures selected to analyze progress in writing in the case of advanced learners, it can be said that the accuracy measures (EFTU/T and Err/T) are the only ones in which there is obvious space for

the level 5 learners to grow from pre to posttest, as illustrated by the difference in performance between these learners and NESs in this area (see section 6.5.1). The fact that not much progress was experienced by the advanced students in fluency, syntactic and lexical complexity may be due to ceiling effects, since the performance demonstrated by these learners was near native-like in the pretest in those areas. Consequently, it can be concluded that either the measures selected (W/T, C/T and Guiraud's Index) were not appropriate to measure progress at the advanced level, or that those areas are not likely to improve once the learners have reached a certain proficiency level. With respect to oral production, it can be claimed that there are also ceiling effects for the advanced students in terms of the measures used to examine syntactic and lexical complexity (C/T and Guiraud's Index), since the performance of the learners on these measures is already native-like in the pretest. Conversely, in fluency (syllables per minute) and accuracy (EFTU/T and Err/T) there was much room for improvement in the pretest, but also in the posttest. Therefore, as was concluded in the case of written production, it can be the case that the measures adopted to analyze oral syntactic and lexical complexity at the advanced level cannot capture the progress experienced by those learners in an English course, or that not much progress is likely to occur in those areas at such a proficiency level.

Another observation which can be made regarding measures is whether they actually measure the language areas they are supposed to assess. As introduced in 5.3.3, there can be a certain controversy about the operationalizations of fluency, syntactic, lexical complexity and accuracy. It was stated that the

measure W/T had been considered to be a measure of productivity or syntactic complexity and not just fluency. The results observed for all the groups analyzed in this dissertation for this measure and the syntactic complexity measure (C/T) are highly comparable, which might mean that, in fact, they could be measuring the same thing or that progress in fluency and syntactic complexity can occur simultaneously.

CHAPTER 8: CONCLUSION

The aim of this dissertation was to shed some light on the issue of the distribution of instruction hours in SLA, analyzing its effect on different language areas for adult learners of different proficiency levels. It was claimed that, whereas most researchers tend to agree that time spent practicing in any skill (whether it is learning an L2 or any other cognitive skill) leads to higher levels of performance, there is no consensus about how the time devoted to such practice should be distributed. The results obtained in the tests included in the present study, which attempted to answer Research Question 1, demonstrate that concentrated L2 instruction has some positive effects in terms of some language skills, namely, listening, grammar, vocabulary, as well as some aspects of written and oral production. Nevertheless, it was also revealed, answering Research Question 2, that the positive effects that can be obtained in intensive language courses are only evident at the intermediate level. The students enrolled in the advanced groups considered for this particular study seem to experience similar language gains from the beginning until the end of their course, irrespective of the type of program which they followed (extensive or intensive).

The intermediate-level students who attended the intensive course (110 hours in one month) showed more language gains than those in the semi-intensive (110 hours in three to four months) and also more than the learners in the extensive program (110 hours in seven months). Moreover, the advantages for intensive instruction were not linear, with more gains being evident in the intensive, then

the semi-intensive and finally the extensive program. While such linearity was usually obtained in the written test that evaluated listening, grammar, vocabulary and reading comprehension, the learners in the extensive program demonstrated more progress than those in the semi-intensive in some measures of written and oral production. It was claimed that one of the positive aspects of intensive language instruction was the disposal of a concentrated amount of time exclusively devoted to the study of an L2 (as was the case for those learners registered in the intensive program). The students registered in the semi-intensive course had to attend English lessons apart from studying the subjects from their degree. Moreover, the time concentration per session is highly different in semi-intensive (2-2.5 hours) and intensive (5 hours) programs. It was also suggested that the differences between the three program types are not pronounced in all the measures under analysis, which may be due to the fact that very few hours of instruction could be considered between the pretest and the posttest (approximately 80 hours).

The advanced students, on the other hand, demonstrated similar improvements in their L2 skills whether they followed an extensive or an intensive course. Some improvement was made in terms of listening, grammar and vocabulary (written test), written fluency and syntactic complexity, and oral fluency. One of the reasons why fewer language gains were made at the advanced level was the fact that ceiling effects existed in some measures of written and oral production, which manifested when comparing the performance of the advanced learners and native English speakers.

The different trends observed for learners at the intermediate and the advanced levels concerning the effect of the distribution of instruction hours on the acquisition of English as a foreign language were explained by the different acquisition stages of the learners in the two groups. Intermediate learners have more room for improvement; consequently, more gains were experienced. Additionally, intensive exposure to input in the L2 benefits the acquisition of frequent patterns, by providing fresh declarative knowledge in a variety of contexts (inside and outside the class) that the learners can easily retrieve and proceduralize. Since the time lapse between sessions is short, the learners can better remember previously heard/seen forms and thus are more ready to retrieve them, and practice them; hence proceduralization is more likely to occur. On the other hand, when the material to be acquired is more complex, less frequent and mostly learned explicitly, such concentrated exposure may not have such positive effects, because the material to be learned is not often present in the outside-of-class input/output or other contexts which foster interaction in intensive programs.

The findings reported in this dissertation are unique in a number of respects. First of all, the design of the study permitted the investigation of time distribution alone, without other confounding variables such as time increase, since the different programs under analysis (extensive, semi-intensive, and intensive) provided the same hours of instruction (110 hours) but distributed differently, in contrast to most studies which have analyzed the effects of intensive instruction (Lightbown & Spada, 1994; Spada & Lightbown, 1989; White & Turner, 2005). Furthermore, this dissertation not only examines data from advanced

learners, who have not received much attention in the program evaluation literature concerning the effect of time distribution on L2 instruction, but also provides comparison data from native-English speakers. Data from L2 learners should be compared, ideally, with data from native speakers in order to examine where students stand with respect to native-like performance and most importantly whether lack of progress is due to ceiling effects.

There are some areas which are left for further research. It was the general aim of this dissertation to examine L2 gains of English learners in intensive and non-intensive programs for all language areas: listening, grammar, vocabulary, reading, writing and speaking. Future research studies could provide a more detailed account for progress in one of these aspects, including more items or more measures in order to analyze more deeply the progress in a particular language skill. Additionally, examining retention after some time has elapsed would be highly valuable in order to observe whether the learners in the intensive program at the intermediate level who show a more advantageous performance at completion of their course still maintain such an advantage with respect to those in the extensive program. Similarly, the study of long-term learning for the advanced students could provide relevant information concerning language gains in a delayed posttest. It might be the case that the differences between the two program types which did not occur at the end of the course could take place after a long period of time. Likewise, a study can be designed in which the same instructors are in charge of different groups receiving intensive and non-intensive instruction.

It must be emphasized that the courses under analyses only provided 110

hours of instruction, and that the hours between the pre and the posttests were approximately 80. It would be useful to analyze progress in the case of students who have received more hours of instruction in the L2 in intensive and non-intensive courses, especially since more hours of instruction may make a difference at the advanced level.

If further research on the effect of time distribution on L2 learning corroborates the results obtained in this particular study, important modifications in the programming of foreign language classes could take place. Such changes could be implemented more easily in language schools or colleges; nevertheless, the implementation of intensive courses in schools could require alterations in the scheduling, which would make intensive language classes of a similar type to the intensive English or intensive French classes in Canada hard to implement.

The distribution of the L2 instruction hours should no longer be "a matter of tradition and guesswork" (Stern, 1985: 18) but rather should start to consider findings from pedagogical and psychological research. More studies should be performed in this area in order to provide evidence for the optimal time distribution for language learning at different proficiency levels. The findings from such studies could lead to a restructuring of language program designs which will aim at making the students' task easier, and foreign language teaching and learning a more efficient process for teachers, students, and institutions providing second language instruction.

REFERENCES

ACT-R 6.0 Tutorial (http://act-r.psy.cmu.edu/actr6/)

ACTFL (1999). *The ACTFL Proficiency Guidelines: Speaking (revised 1999).* Yonders, NY: ACTFL.

Agard, F. B. (1946). The University of Chicago Language Investigation Tests. *Hispania, 29*, 31-37.

Agard, F. B., Clements, R. J., Hendrix, W. S., Hocking E., Pitcher, S. L., van Eerden, A., & Doyle, H. G. (1945). A survey of language classes in the Army Specialized Traning Program. *The Modern Language Journal, 29*, 155-160.

Anderson, J. R. (1976). *Language, Memory, and Thought.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. (1980). *Cognitive Psychology and Its Implications.* New York: W. H. Freeman and Company.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

Anderson, J. R. (1983). *The Architecture of Cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1993). *Rules of the Mind.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111,* 1036-1060.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79*, 97-123.

Anderson, J. R., & Finchman, J. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1322-1340.

Anderson, J. R., Finchman, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 932-945.

Appleton, S. L., Bjork, R. A., & Wickens, T. D. (2005). Examining the spacing effect in advertising: Encoding variability, retrieval processes and their interaction. *Journal of Consumer Research*, 32, 266-276.

Arnaud, P. L. J. (1992). Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate-component tests. In P. J. L. Arnaud & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 133-145). London: Macmillan.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory Vol. 2* (pp. 89-195). New York: Academic Press.

Austin, S. D. M. (1921). A study in logical memory. *American Journal of Psychology, 32,* 370-403.

Austin, T. L., Fennel, R. R., & Yeager, C. R. (1988). Class scheduling and academic achievement in a non-traditional graduate program. *Innovative Higher Education, 12,* 79-90.

Baddeley, A. (1990). *Human Memory: Theory and Practice.* Needham Heights, MA: Allyn and Bacon.

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology, 108*, 296-308.

Bahrick H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 344-349.

Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition, 11*, 17-34.

Benseler, D., & Schulz, R. A. (1979). Intensive foreign language courses. *Language in Education: Theory and Practice, 18*, 1-53.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition, 16,* 157-168.

Billy, L. (1980). Expérimentation d'une nouvelle approche en immersion. *Canadian Modern Language Review, 36*, 422-33.

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. S. Sykes (Eds.), *Practical Aspects of memory : Current Research and Issues* (pp. 396-401). Chichester: Wiley.

Bloom, K. C., & Shuell T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research, 74*, 245-248.

Brann, H. W. (1944). The conversational method in modern language teaching: Its advantages and limitations. *The German Quarterly, 17*, 209-215.

Braun, K., & Rubin, D. C. (1998). The spacing effect depends on an encoding deficit, retrieval and time in working memory: Evidence from once-presented words. *Memory, 6*, 37-65.

Brecht, R. D., Davidson, D. E., & Ginsberg, R. B. (1995). Predictors of foreign language gain during study abroad. In B. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context* (pp. 37-66). Amsterdam/Philadelphia: John Benjamins.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing* (pp. 23-48). Harlow, England: Longman.

Byrd, C. W. (1980). Intensive language instruction at a small Liberal Arts College: The Dartmouth Model at Emory & Henry. *The Modern Language Journal*, *64*, 297-30.

Carlson, R. A., Khoo, B., Yaure, R. G., & Schneider, W. (1990). Acquisition of problem-solving skill: Levels of organization and use of working memory. *Journal of Experimental Psychology: General, 119,* 193-214.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64,* 723-733.

Carroll, J. B. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals, 1*, 131–51.

Carroll, J. B. (1974). Learning theory for the classroom teacher. In G. Jarvis (Ed.), *The Challenge of Communication* (*ACTFL Review of Foreign Language Education Vol. 6*) (pp. 329-351). Skokie, IL: National Textbooks.

Carroll, J. B. (1981). Conscious and automatic processes in language learning. *The Canadian Modern Language Review, 37*, 462–64.

Carroll, J. B. (1989). The Carroll Model: A 25-year retrospective and prospective view. *Educational Researcher, 18,* 26-31.

Carroll, J. B., & Sapon, S. S. (1955). *Modern Language Aptitude Test*. New York: Psychological Corporation.

Carroll, J. M. (1987). The Copernican Plan: A concept paper for restructuring high schools. Paper presented at the Annual Meeting of the American Association of School Administrators (New Orleans, LA, February 20-23, 1987).

Carroll, J. M. (1994). The Copernican Plan evaluated. *Phi Delta Kappan, 76,* 104-113.

Casanave, C. P. (1994). Language development in students´ journals. *Journal of Second Language Writing, 3*, 179-201.

Caskey, S. R. (1995). Learning outcomes in intensive courses. *Journal of Continuing Higher Education, 42,* 23-27.

Celaya, M. L., Pérez-Vidal, C., & Torras, M. R. (2001). Matriz de criterios de

medición para la determinación del perfil de competencia lingüística escrita en inglés. *RESLA, 14,* 87-98.

Celaya, M. L., Torras, M. R., & Pérez-Vidal, C. (2001). Short and mid- term effects of an early start: An analysis of EFL written production. *Eurosla Yearbook, 1*, 195-209.

Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 19,* 389-396.

Cheng, P. W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. *Psychological review, 92,* 214-23.

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs and conventional actions from massed or distributed exposures. *Developmental Psychology, 38*, 967-978.

Cipolla, W. C. (1982). To speak or not to speak: Promoting the Dartmouth Model at a small private college. *ADFL Bulletin*, *13*, 6-7.

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition, 26,* 227-248.

Collentine, J., & Freed, B. (2004). Learning context and its effects on second language acquisition: Introduction. *Studies in Second Language Acquisition*, *26*, 153-171.

Collins, L., Halter, R. H., Lightbown, P. M. & Spada, N. (1999). Time and the distribution of time in L2 instruction. *TESOL Quarterly, 33*, 655-680.

Cooper, E. C., & Pantle, A. J. (1967). The total time hypothesis in verbal learning. *Psychological Bulletin, 68,* 221-234.

Cooper, T. C. (1976). Measuring written syntactic patterns of second language learners of German. *The Journal of Educational Research, 69,* 176-183.

Council of Europe (2001). *Common European Framework of Reference for Languages.* Cambridge: CUP.

Cowan J. M., & Graves, M. (1944). A statement of intensive language instruction. *The German Quarterly, 17*, 165-166.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition, 11,* 367-383.

Cuddy, L. J., & Jacoby L. J. (1982). When forgetting helps memory: An analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior, 21*, 451-467.

Currall, S., & Kirk, R. E. (1986). Predicting success in intensive language courses. *The Modern Language Journal, 70,* 107-113.

Cushman, K. (1989). Schedules that bind. *American Educator, 13*, 35-59.

Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in spontaneous speech of bilinguals. *Applied Linguistics, 24,* 197-222.

Daniel, E. L. (2000). A review of time-shortened courses across disciplines. *College Student Journal, 34,* 298-308.

DeKeyser, R. M. (1991). Foreign language development during a semester abroad. In B. F. Freed (Ed.), *Foreign Language Acquisition and the Classroom* (pp. 104-118). Lexington, MA: D. C. Heath.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19*, 195-221.

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 125-151). Cambridge: CUP.

DeKeyser, R. M. (2007a). Introduction: Situating the concept of practice. In R. M. DeKeyser (Ed.), *Practice in Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 1-18). Cambridge: CUP.

DeKeyser, R. M. (2007b). Study abroad as foreign language practice. In R. M. DeKeyser (Ed.), *Practice in Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 208-226). Cambridge: CUP.

DeKeyser, R. M. (2007c). *Practice in Second Language: Perspectives from Applied Linguistics and Cognitive Psychology.* Cambridge: CUP.

Delaney, P. F., & Knowles, M. E. (2005). Encoding strategy changes and spacing effects in the free recall of unmixed lists. *Journal of Memory and Language, 52*, 120-130.

Dellarosa, D., & Bourne, L. E. (1985). Surface form and the spacing effect. *Memory & Cognition, 13*, 529-537.

de López, M. (1996). Block scheduling and second laguage instruction. *ACTFL Newsletter, 12*, 11-15.

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79*, 162-170.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist, 43*, 627-634.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. Ligon & R. Bjork (Eds.), *Memory* (pp. 317-344). San Diego: Academic Press.

Deveny, J. J., & Bookout, J. C. (1976). The intensive language course: Towards a successful approach. *Foreign Language Annals, 9,* 58-63.

Dewey, D. P. (2004). A comparison of reading development by learners of Japanese in intensive domestic immersion and study abroad contexts. *Studies in Second Language Acquisition, 16*, 303-327.

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition, 26*, 249-273.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84*, 795-805.

Ebbinghaus, H. (1913). *Memory* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Teachers College. (Original work published 1885). (Paperback ed., New York: Dover, 1964).

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18*, 91-126.

Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 33-68). Cambridge: CUP.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*, 143-188.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*, 305-352.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning, 54*, 227-275.

Elmes, D. G., Craig, J. D., & Herdelin, N. J. (1983). What is the role of affect in the spacing effect? *Memory & Cognition, 11*, 144-151.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*, 725-747.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychologial Review, 100*, 363-406.

Fillmore, C. (1979). On fluency. In C. Fillmore, D. Kempler, & W. Wang (Eds.), *Individual Differences in Language Ability and Language Behavior* (pp. 85-101). New York: Academic Press.

Folse, K. (2004). *Vocabulary Myths.* Michigan: University of Michigan Press.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*, 299-323.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken discourse: A unit for all reasons. *Applied Linguistics, 21*, 333-353.

Frank, T. (1972). An intensive program in German. *ADFL Bulletin, 3*, 31-32.

Freed, B. F. (1990). Language learning in a study abroad context: The effects of interactive and non-interactive out-of-class contact on grammatical achievement and oral proficiency. In J. E. Alatis (Ed.), *Linguistics, Language Teaching and Language Acquisition: The Interdependence of Theory, Practice and Research* (GURT 1990) (pp. 459-477). Washington, D. C.: Georgetown University Press.

Freed, B. F. (1995). What makes us think that students who study abroad become fluent? In B. F. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context* (pp. 123-148). Amsterdam/Philadelphia: John Benjamins.

Freed, B. F. (1998). An overview of issues and research in language learning in a study abroad setting. *Frontiers, 4*, 31-60.

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, *26*, 275-301.

Fuerst, N. (1944). Towards defining the intensive course. *The German Quarterly, 17*, 201-204.

Gardner, R. C., Smythe, P. C., & Brunet, G. R. (1977). Intensive second language study: Effects on attitudes, motivation and French achievement. *Language Learning, 27*, 243-261.

Gass, S., Mackey, A., Álvarez-Torres, M. J., Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning, 49*, 4, 549-581.

Gay, L. R. (1974). Temporal position of reviews and its effect on the retention of mathematical rules. *Journal of Educational Psychology, 64*, 171-182.

Genesee, F. (1987). *Learning through Two Languages: Studies of Immersion and Bilingual Education.* Rowley: Newbury House.

Genesee, F. (1994). *Integrating Language and Content: Lessons from Immersion.* The National Center for Research on Cultural Diversity and Second Language Learning, Educational Practice Report 11, 1-15.

Genesee, F. (2004). *What do we know about bilingual education for majority language students?* In T. K. Bhatia, & W. Ritchie (Eds.), *Handbook of Bilingualism and Multiculturalism* (pp. 547-576). Malden, MA: Blackwell.

Germain, C., Lightbown, P., Netten, J., & Spada, N. (2004). Intensive French and Intensive English: Similarities and differences. *The Canadian Modern Language Review, 60*, 409-430.

Germain, C., Netten, J., & Movassat, P. (2004). L'évaluation de la production orale en français intensif: critères et résultats. *The Canadian Modern*

Language Review, 60, 309-332.

Germain, C., Netten, J., & Séguin, S. P. (2004). L'évaluation de la production écrite en français intensif: critères et resultats. *The Canadian Modern Language Review, 60*, 333-354.

Gilabert, R. (2005). *Task Complexity and L2 Narrative Oral Production*. Ph.D. Dissertation, Universitat de Barcelona.

Girard, D. P. (1943). The teaching of foreign languages during and after the war. *The French Review, 17*, 7.

Glanzer, M., & Duarte, A. (1971). Repetition between and within languages in free recall. *Journal of Verbal Learning and Verbal Behavior, 10*, 625-630.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7*, 95-112.

Glover, J. A., & Corkill, A. J. (1987). Influence of paraphrased repetitions on the spacing effect. *Journal of Educational Psychology, 79*, 198-199.

Goedsche, C. R. (1946). The semi-intensive course at Northwestern. *The German Quarterly, 19*, 42-47.

Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology, Memory and Cognition, 15*, 371-377.

Greeno, J. G. (1970). Conservation of information-processing capacity in paired-associate memorizing. *Journal of Verbal Learning and Verbal Behavior, 9*, 581-586.

Griffiths, R. (1991). Pausological research in an L2 context: A rationale, and review of selected studies. *Applied Linguistics, 12*, 345-364.

Hills, E. C. (1919). Has the War proved that our methods of teaching modern languages in the colleges are wrong? *The Modern Language Journal, 4*, 1-13.

Hinger, B. (2006). The distribution of instructional time and its effect on group cohesion in the foreign language classroom: A comparison of intensive and standard format courses. *System, 34*, 97-118.

Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *The Psychology of Learning and Memory* (pp. 47-91). New York: Academic Press.

Hintzman, D. L., Summers, J. J., & Block, R. A. (1975). What causes the spacing effect? Some effects of repetition, duration, and spacing on memory for pictures. *Memory & Cognition, 3*, 287-294.

Hirsh, B. B. (1982). Beyond Rassias: Intensive language programs at large and small institutions. *ADFL Bulletin, 13*, 1-2.

Homburg, T. J. (1984). Holistic evaluation of ESL compositions: Can it be validated objectively? *TESOL Quarterly, 18*, 87-107.

Howard, E. R. (2003). *Biliteracy development in two-way immersion education programs: A multilevel analysis of the effects of native language and home language use on the development of narrative writing ability in English and*

Spanish. Unpublished Ph.D. Dissertation, Harvard University.

Huey, E. B. (1968). *The Psychology and Pedagogy of Reading.* Cambridge, MA: MIT Press.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 258-287). Cambridge: CUP.

Hulstijn, J. H. (2002). Towards a unified account of the representation, processing and language acquisition of second language knowledge. *Second Language Research, 18,* 193-223.

Hunt, K. W. (1965). *Differences in Grammatical Structures Written at Three Grade Levels. Research Report No. 3*. Urbana, IL: National Council of Teachers of English.

Ife, A., Vives-Boix, G., & Meara, P. (2000). The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics, 4*, 55-84.

Irmsher, K. (1996). Block Scheduling. *ERIC Digest, 104.*

Ittner, R. L. (1944). Implications of the Armed Forces´ language program. *The German Quarterly, 17*, 176-182.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17,* 649-667.

Johnson, K. (1996). *Language Learning & Skill Learning*. Oxford: Blackwell.

Johnson, K. (2001). *An Introduction to Foreign Language Learning and Teaching*. Harlow: Longman.

Jurasek, R. (1982). Intensive and super German at Earlham college. *ADFL Bulletin, 13,* 3-5.

Kahana, M. J., & Howard, M. W. (2005). Spacing and lag effects in free recall of pure lists. *Psychonomic Bulletin & Review*, *12*, 159-164.

Kalivoda, T. B. (1975). Organization of intensive instruction: Dispelling misconceptions and facing realities. *Hispania, 58*, 114-121.

Kasworm, C. E. (2003). From the adult student's perspective: Accelerated degree programs. *New Directions for Adult and Continuing Education, 97,* 17-27

Keilstrup, D. V. (1981). Practical guidelines and activities for an advanced foreign language intensive program. *The Modern Language Journal, 65*, 377-82.

Kilker, J. A., & Gunderson, F. (1972). Intensely French: A new program. *The French Review, 45*, 850-854.

Krug D., Brandon Davis, T., & Glover, J. A. (1990). Massed versus distributed repeated reading: A case of forgetting helping recall? *Journal of Educational*

Psychology, 82, 366-371.

Lafford, B. A. (2004). The effect of the context of learning on the use of communication strategies by learners of Spanish as a second language. *Studies in Second Language Acquisition, 26*, 201-225.

Lambert, W., & Tucker, G. R. (1972). *Bilingual Education of Children: The St. Lambert Expermiment*. Cambridge, MA: Newbury House.

Landauer, T. K. (1969). Reinforcement as consolidation. *Psychological Review, 76*, 82-96.

Lapkin, S., Hart, D., & Harley G. (1998). Case study of compact French models: Attitudes and achievement. In S. Lapkin (Ed.), *French Second Language Education in Canada* (pp. 31-55). Toronto: University of Toronto Press.

Lapkin, S., Hart, D., & Swain, M. (1991). Early and middle French immersion programs: French language outcomes. *The Canadian Modern Language Review, 47*, 635-641.

Lapkin, S., Hart, D., & Swain, M. (1995). A Canadian interprovincial exchange: Evaluating the linguistic impact of a three-month stay in Quebec. In B. F. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context.* Amsterdam/Philadelphia: John Benjamins.

Larsen-Freeman, D. (1983). Assessing global second language proficiency. In H. W. Seliger & M. Long (Eds.), *Classroom-Oriented Research in Second Language Acquisition* (pp. 287-304). Rowley, MA: Newbury House.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese Learners. *Applied Linguistics, 27,* 590-619.

Larsen-Freeman, D., & Strom, V. (1977). The construction of a second language acquisition index of development. *Language Learning, 27*, 123-134.

Leavitt, S. E. (1943). Why waste time? Report on the preliminary Spanish Program, University of North Carolina. *Hispania, 26*, 310-311.

Leeman, J. (2007). Feedback in L2 learning: Responding to errors during practice. In R. M. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 111-138). Cambridge: CUP.

Lenneberg, E. H. (1967). *Biological Foundations of Language.* New York: Wiley.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40,* 387-417.

Leow, R. P. (2007). Input in the L2 classroom: An attentional perspective on receptive practice. In R. M. DeKeyser (Ed.), *Practice in Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 21-50). Cambridge: CUP.

Levelt, W. (1989). *Speaking: From Intention to Articulation.* Cambridge, MA: MIT Press.

Levelt, W. (1993). Language use in normal speakers and its disorders. In G. Blanken, J. Dittman, H. Grimm, J. Marshall, & C. Wallesch (Eds.), *Linguistic Disorders and Pathologies: An International Handbook* (pp. 1-15). Berlin: de Gruyter.

Lewis, C. W., Winokur, M. A., Cobb, R. B., Gliner, G. S., & Schmidt, J. (2005). *Block Scheduling in the High School Setting: A Synthesis of Evidence-Based Research*. Berkeley, CA: MRP Associates.

Lightbown, P. M., & Spada, N. (1991). Étude des effets à long terme de l'apprentissage intensif de l´anglais, langue seconde, au primaire. *The Canadian Modern Language Review, 48*, 90-117.

Lightbown, P. M., & Spada, N. (1994). An innovative program for primary ESL students in Quebec. *TESOL Quarterly, 28*, 563-579.

Lightbown, P. M., & Spada, N. (1997). Learning English as a second language in a special school in Quebec. *The Canadian Modern Language Review, 53,* 315-355.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492-527.

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12,* 251–285.

Luplow, C. A. (1982). The Rassian Method and the teaching of Russian. *The Slavic and East European Journal, 26*, 216-226.

MacIntyre, P. D., Baker, S. C., Clément, R., & Donovan L. A. (2003). Talking in Order To Learn: Willingness To Communicate and Intensive Language Programs. *The Canadian Modern Language Review, 59,* 589-607.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy tradeoff in skilled behavior. *Psychological Review, 89,* 483-506.

Mackey, A. (2007). Interaction as practice. In R. M. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 85-110). Cambridge: CUP.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical Diversity and Language Development.* New York: Palgrave Macmillan.

Mammarella, N., Russo, R., & Avons, S. E. (2002). Spacing effects in cued-memory tasks for unfamiliar faces and nonwords. *Memory & Cognition, 30,* 1238-1251.

Marti, G. (1972). An intensive French experience: Some observations and results. *The French Review, 14*, 1145-1151.

McFarland, C. E., Rhodes, D. D., & Frey, T. J. (1979). Semantic-feature variability and the spacing effect. *Journal of Verbal Learning and Verbal Behavior, 18,* 163-172.

McGeoch, J. A. (1943). *The Psychology of Human Learning.* New York: Longmans Green.

McKee, E. (1983). The effects of intensive language instruction on student performance in beginning college French. Paper presented at the Central States Conference on the Teaching of Foreign Languages (St. Louis, MO, March 24-26, 1983).

McLaughlin, B. (1987). *Theories of Second Language Learning*. London: Edward Arnold.

McLaughlin, B. (1990). Restructuring. *Applied Linguistics, 11,* 113-128.

McMullen, E. (1950). The intensive method: An experiment. *The French Review, 23*, 316-318.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition, 20*, 52-83.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 596-606.

Menut, A. D. (1953). Intensive language courses: Content and techniques. *The Modern Language Journal*, *37,* 189-194.

Metcalf, G. (1946). Experiments with intensive language teaching at the University of Chicago. *The German Quarterly, 19*, 6-8.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Milton, J., & Meara, P. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL Review of Applied Linguistics, 107-108*, 17-34.

Ministerio de Educación y Ciencia, INECSE (2004). Evaluación de la enseñanza y el aprendizaje de la lengua inglesa: Educación Secundaria Obligatoria 2001: Informe final. Madrid: Secretaría General de Educación.

Ministerio de Educación y Cultura, INCE (2000). Evaluación de la enseñanza y el aprendizaje de la lengua inglesa: Educación primaria 1999: Avance de resultados. Madrid: Centro de Publicaciones del Ministerio de Educación y Cultura.

Miralpeix, I. (2006). Age and vocabulary acquisition in English as a Foreign Language (EFL). In C. Muñoz (Ed.), *Age and the Rate of Foreign Language*

Learning (pp. 89-106). Clevedon: Multilingual Matters.

Möhle, D. (1984). A comparison of the second language speech production of different native speakers. In H. Dechert, D. Möhle, & M. Raupach (Eds.), *Second Language Productions* (pp. 26-49). Tübingen: Gunter Narr.

Moor, P., & Cunningham, S. (2003). *Cutting Edge Advanced.* Harlow, England: Longman

Moore, M. (1950). Capsule Courses. *The French Review, 33*, 496-500.

Moreland, F. L. (1980). An Intensive Approach to Latin and Greek: The Latin/Greek Institute of the City University of New York. *Classical Outlook, 58*, 5-8.

Muñoz, C. (2004). A comparison of students' achievement in English in regular programmes and in CLIL programmes. Unpublished manuscript. UB.

Muñoz, C. (2006). The Effects of age on foreign language learning: the BAF project. In C. Muñoz (Ed.), *Age and the Rate of Foreign Language Learning* (pp. 1-40). Clevedon: Multilingual Matters.

Muñoz, C. (2007). Age-related differences and second language learning practice. In R. M. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 229-255). Cambridge: CUP.

Muranoi, H. (2007). Output practice in the L2 classroom. In R. M. DeKeyser (Ed.), *Practice in Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 51-84). Cambridge: CUP.

National Education Commission on Time and Learning (1994). *Prisoners of time.* Washington, D.C.: U.S. Government Printing Office.

Navés, T. (2006). *The Long-term Effects of an Early Start on Writing in Foreign Language Learning.* Unpublished Ph.D. Dissertation, Universitat de Barcelona.

Netten, J. & Germain, C. (2000). Transdisciplinary approach and intensity in second language learning/teaching. *The Canadian Journal of Applied Linguistics, 3*, 107-122.

Netten, J., & Germain, C. (2004a). Intensive French: Introduction. *The Canadian Modern language Review, 60,* 263-273.

Netten, J., & Germain, C. (2004b). Theoretical and research foundations of intensive French. *The Canadian Modern Language Review, 60,* 275-294.

Netten, J., & Germain, C. (2005). *Literacy in the Canadian Context and Intensif French / La littératie dans le contexte canadien et le français intensif*. Paper presented at the Conference on Literacy and Numeracy, organized by the

Literacy and Numeracy Secretariat of the Ontario Ministry of Education with the support of the Council of Ministers of Education, Canada. Toronto, November 2005.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive Skills and their Acquisition* (pp. 1-55). Hillsdale, NJ: Erlbaum

Nixon, R. O. (1996). A source document on accelerated courses and programs at accredited two & four year colleges and universities. ERIC Document Reproduction Service No. ED 399 827.

Nordsieck, R. (1946). A brief report on intensive, semi-intensive, and special courses in German at the Ohio State University. *The German Quarterly, 19*, 48-51.

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review, 75,* 522-536.

O'Donnell, R. C. (1976). A critique of some indices of syntactic maturity. *Research in the Teaching of English, 10*, 31-38.

O'Malley, J. M. & Chamot, A. U. (1990). *Learning Strategies in Second Language Acquisition.* Cambridge: CUP.

Ortega, L. (2007). Meaningful L2 practice in foreign language classrooms: A cognitive-interactionist SLA perspective. In R. M. DeKeyser (Ed.), *Practice in Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 180-207). Cambridge: CUP.

Oswald, V. A., Vail, C. C. D., Goedsche, C. R., & Ulmer, B. (1950). Procedures and results in intensive courses. *The German Quarterly, 23,* 129-150.

Oxenden, C., & Latham-Koenig, C. (2001). *English File Intermediate.* Oxford: OUP.

Padín, J. (1945). Tendencias recientes en la enseñanza de lenguas vivas. *Hispania, 28*, 61-68.

Pargment, M. S. (1945). On learning a foreign language. *The Modern Language Journal, 29*, 198-209.

Paulsen, W. (1944). The A.S.T.P. experiment and our future language courses. *The German Quarterly, 17*, 167-175.

Pavlik, P. I. (2003). An ACT-R model for the spacing effect. *ICCM, 5* 177-182.

Pavlik, P. I. (2005). *The microeconomics of learning: Optimizing paired-associate memory*. Unpublished Ph.D. Dissertation, Carnegie Mellon University.

Pavlik, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science, 29,* 559-586.

Pérez-Vidal, C., Torras, R., & Celaya, M. L. (2000). Age and EFL written performance by Catalan / Spanish bilinguals. *Spanish Applied Linguistics, 4*, 267-290.

Petadella, R. A. (1988). DILM at Baruch College: An open letter. *Hispania, 68*, 199-201

Peters, M. (2000). Le bain linguistique: Une innovation pour le programme de français de base. *Apprentissage et Socialisation, 20*, 71-86.

Polanyi, L. (1995). Language learning and living abroad: Stories from the field. In B. F. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context* (pp. 271-91). Amsterdam/Philadelphia: John Benjamins.

Polio, C., Fleck, C., & Leder, N. (1998). "If I only had more time": ESL learners changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing, 7*, 43-68.

Postman, L., & Knecht, K. (1983). Encoding variability and retention. *Journal of Verbal Learning and Verbal Behavior, 22*, 133-152.

Queen, J. A. (2000). Block Scheduling Revisited. *Phi Delta Kappan, 82*, 214-221.

Raaijmakers, J. G. W. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science, 27*, 431-452.

Ranta, L., & Lyster, R. (2007). A cognitive approach to improving immersion students' oral language abilities: The awareness-practice-feedback sequence. In R. M. DeKeyser (Ed.), *Practice in Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 141-160). Cambridge: CUP.

Regan, V. (1995). The acquisition of sociolinguistic native speech norms: Effects of a year abroad on second language learners of French. In B. F. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context* (pp. 245-267). Amsterdam/Philadelphia: John Benjamins.

Regan, V. (1998). Sociolinguistics and language learning in a study abroad context. *Frontiers, 4*, 61-90.

Riches, N. G., Tomasello, M., & Conti-Ramsden, G. (2005). Verb learning in children with SLI: Frequency and spacing effects. *Journal of Speech, Language, and Hearing Research, 48*, 1397-1411.

Robinson, P. (1995). Attention, memory, and the 'noticing' hypothesis. *Language Learning, 45*, 283-331.

Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced and instructed conditions. *Studies in Second Language Acquisition, 19*, 223-247.

Robinson, P. (2001). Task complexity, cognitive resources, and syllabus

design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 287-318). Cambridge: CUP.

Robinson, P. (2003a). Attention and memory. In C. Doughty & M. H. Long (Eds.), *Handbook of Second Language Acquisition* (pp. 631-678). Oxford: Blackwell.

Robinson, P. (2003b). The Cognition Hypothesis, task design, and adult task-based language learning. *Second Language Studies, 21*, 45-105.

Robinson, P. (2007). Aptitudes, abilities, contexts and practice. In R. DeKeyser (Ed.), *Practice in Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 256-286). Cambridge: CUP.

Rogers, P. P. (1945). Lessons of the ASTP of language teaching for normal times. *Hispania, 28*, 44-49.

Rosenbloom, P., & Newell, A. (1987). Learning by chunking: A production system model of practice. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production System Models of Learning and Development* (pp. 221-286). Cambridge, MA: MIT Press.

Ross, B. H., & Landauer, T. K. (1978). Memory for at least one of two items: Test and failure of several theories of spacing effects. *Journal of Verbal Learning and Verbal Behavior, 17,* 669-680.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89,* 63-77.

Russo, R., & Mammarella, N. (2002). Spacing effects in recognition memory: When meaning matters. *European Journal of Cognitive Psychology, 14*, 49-59.

Russo, N., Mammarella N., & Avons, S. E. (2002). Toward a unified account of spacing effects in explicit cued-memory tasks. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 28,* 819-829.

Sacks, N. P. (1967). "Modern Spanish" in an intensive program for graduate students: An experiment and some reflections. *Hispania, 1,* 310-319.

Santosuosso, J. J. (1957). ASTP characteristics in 55 colleges 1941-1951: Problems of application. *The Modern Language Journal, 41*, 9-14.

Sato, C. (1988). Origins of complex syntax in interlanguage development. *Studies in Second Language Acquisition, 10*, 371–395.

Schachter, J. (1993). Second language acquisition: Perceptions and possibilities. *Second Language Research, 9*, 2, 173-187.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14,* 357-385.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second*

Language Instruction (pp. 3-32). Cambridge: CUP.

Schneider, G. K. (1977). A model for a successful summer foreign language program in German. *ADFL Bulletin, 8*, 27-29.

Schneider, V. I., Healy, A. F., & Bourne, L. E. (1998). Contextual interference effects in foreign language vocabulary acquisition and retention. In A. F. Healy & L. E. Bourne (Eds.), *Foreign Language Learning: Psycholinguistic Studies on Training and Retention* (pp. 77-90). Mahwah, NJ: Lawrence Associates.

Schneider, V. I., Healy, A. F., & Bourne, L. E. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention and transfer. *Journal of Memory and Language, 46,* 419-440.

Schueler, H. (1944). Foreign language teaching under the Army Specialized Training Program. *The German Quarterly, 17,* 183-191.

Schulz, R. A. (1979). Intensive language instruction: How and where it works. *ADFL Bulletin, 11*, 37-41.

Scott, P. A. (1995). Learning experiences in intensive and semester-length classes: Students' voices and experiences. *College Student Journal, 29*, 207-213.

Scott, P. A. (1996). Attributes of a high quality intensive course. *College Student Journal*, *30,* 69-77.

Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology, 19,* 107-122.

Seamon, M. (2004). Short- and long-term differences in instructional effectiveness between intensive and semester-length courses. *Teachers College Record, 106*, 852-874.

Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (Ed.), *Perspectives on Fluency* (pp. 200-219). Ann Arbor: University of Michigan Press.

Segalowitz, N. (2003). Automaticity and second languages. In C. Doughty & M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 382-408). Oxford: Blackwell Publishers.

Segalowitz, N. (2007). Access fluidity, attention control, and the acquisition of fluency in a second language. *TESOL Quarterly, 41,* 181-186.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in At Home and Study Abroad contexts. *Studies in Second Language Acquisition, 26*, 173-199.

Segalowitz, N., & Segalowitz, S. J. (1993). Skilled performance, practice, and

the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics, 14,* 369-385.

Serrano, R., & Howard, E. R. (2007). Second language writing development in English and in Spanish in a two-way immersion program. *International Journal of Bilingualism and Bilingual Education, 10,* 152-171.

Serrano, R., & Muñoz, C. (2007). Same hours, different time distribution: Any difference in EFL? *System, 35,* 305-321.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127-190.

Shuqiang, Z. (1989). Cognitive complexity and written production in English as a second language. *Language Learning*, *37*, 469-481.

Singleton, D. (2005). The critical period hypothesis: A coat of many colours. *International Review of Applied Linguistics in Language Teaching*, *43*, 269-285.

Skehan, P. (1986). The role of foreign language aptitude in a model of school learning. *Language Testing, 3,* 188-221.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*, 185-211.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 183-205). Cambridge: CUP.

Soars, J., & Soars, L. (2005). *New Headway Advanced.* Oxford: OUP.

Spada, N., & Lightbown, P. M. (1989). Intensive ESL programs in Quebec primary schools. *TESL Canada Journal, 7*, 11-32.

Spieler, D. H., & Balota, D. A. (1996). Characteristics of associative learning in younger and older adults: Evidence from an episodic priming paradigm. *Psychology and Aging, 11*, 607-620.

Springer, O. (1944). Intensive language study as a part of the college curriculum. *The German Quarterly, 17*, 224-240.

SPSS 11.5.1 for Windows [Computer software] (2002). SPSS Inc.

St. John, M. F., & Gernbacher, M. A. (1998). Learning and losing syntax: Practice makes perfect and frequency builds fortitude. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign Language Learning: Psycholinguistic Studies on Training and Retention* (pp. 231-253). Mahwah, NJ: Lawrence Associates.

Stansfield, C., & Honor, J. (1981). The Dartmouth-Rassias Model of teaching foreign languages. *ADFL Bulletin*, *12*, 23-27.

Stephens, D. T. (1978). An illustration of "motivating students in the foreign

language classroom" in an intensive French course. *Foreign Language Annals, 11,* 25-29.

Stern, H. H. (1985). The time factor and compact course development. *TESL Canada Journal / Revue TESL du Canada, 3,* 12-27.

Swain, M. (1981). Time and timing in bilingual education. *Language Learning, 31,* 1-15.

Swain, M. (1985). Bilingual education for the English-speaking Canadian. In J. E. Alatis & J. J. Staczek (Eds.), *Perspectives on Bilingualism and Bilingual Education* (pp. 385-398). Washingon, D.C.: Georgetown University Press.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 64-81). Cambridge: CUP.

Swain, M., & Barik, H. C. (1976). *Five Years of French Immersion*. Toronto, Ontario: The Ontario Institute for Studies in Education.

Swain, M., & Lapkin, S. (1986). Immersion French at the secondary level: "The goods" and "the bads". *Contact, 5,* 2-9.

Tamarkin, T. (1988). Intensive language programs: Two models for the community college international language program. *Hispania, 71,* 177-179.

Thios S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior, 15,* 529-536.

Thomas, M. (1994). Assessment of L2 proficiency in second language acquisition research. *Language Learning, 44,* 307-336.

Thorndike, E. L. (1916). Notes on practice, improvability and the curve of work. *American Journal of Psychology, 27,* 550-565.

Toppino, T. C. (1991). The spacing effect in young children's free recall: Support for automatic process explanations. *Memory & Cognition, 19,* 159-167.

Toppino T. C., & Bloom, L. C. (2002). The spacing effect, free recall, and two-process theory: A closer look. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28,* 437-444.

Toppino, T. C., & DeMesquita, M. (1984). Effects of spacing repetitions on children's memory. *Journal of Experimental Child Psychology, 37,* 637-648.

Toppino, T. C., & DiGeorge, W. (1984). The spacing effect in free recall emerges with development. *Memory & Cognition, 12,* 118-122.

Toppino, T. C., Hara, Y., & Hackman, J. (2002). The spacing effect in the free recall of homogeneous lists: Present and accounted for. *Memory & Cognition, 30,* 601-606.

Torras, M. R., Navés, T., Celaya, M. L., & Pérez-Vidal, C. (2006). Age and IL

Development in Writing. In C. Muñoz (Ed.), *Age and the Rate of Foreign Language Learning* (pp. 156-182). Clevedon: Multilingual Matters.

Towell, R., & Hawkins, R. (1994). *Approaches to Second Language Acquisition.* Clevedon: Multilingual Matters.

Turnbull, M., Lapkin, S., Hart, D., & Swain, M. (1998). Time on task and immersion graduates' French proficiency. In S. Lapkin (Ed.), *French Second Language Education in Canada* (pp. 31-55). Toronto: University of Toronto Press.

Ullman, M. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research, 30*, 37-69.

Underwood, B. J. (1961). Ten years of massed practice on distributed practice. *Psychological review, 68*, 229-247.

Urbanski, H. (1982). Summer and weekend immersion programs at New Paltz. *ADFL Bulletin, 13*, 10-11.

Van Mondfrans, A. P., Schott, J. L., & French, D. G. (1972). Comparing block scheduling and traditional scheduling on student achievement and attitudes. Paper presented at the annual convention of the American Educational Research Association (Chicago, IL, April 1972).

VanPatten, B. (1990). Attending to form and content in the input : An experiment in consciousness. *Studies in Second Language Acquisition, 12*, 287-301.

Van Syoc, L. J., & Gleason, J. (1993). Traditional or intensive courses lengths? A comparison of outcomes in economics learning. *Journal of Economics Education, 24*, 15-22.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 796-800.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2005). Limitations to the spacing effect: Demonstration of an inverted u-shaped relationship between interrepetition spacing and free recall. *Experimental Psychology, 52*, 257-263.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing, 17*, 65-83.

Wallace, J. E. (1972). Three weeks equals thirty weeks? A report on a experimental intensive January language course. *Foreign Language Annals, 6*, 88-94.

Wallinger, L. M. (2000). The effect of block scheduling on foreign language learning. *Foreign Language Annals, 33*, 36-50.

Watson, C. (1998). Instructional ideas for teaching in block schedules. *Kappa Delta Pi Record, 34,* 94-97.

Wesche, M., MacFarlane, A., & Peters, M. (1994). *The bain linguistique: A core French experiment at Churchill alternative school, 1993-1994.* Final report to the Ottawa Board of Education, Ottawa, Canada.

White, J., & Turner, C. E. (2005). Comparing children's oral ability in two ESL programs. *The Canadian Modern Language Review, 61,* 491-517.

Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior, 16,* 465-478.

Wickens, C. D. (1989). Attention and skilled performance. In D. H. Holding (Ed.), *Humans Skills* (pp. 71-105). New York: John Wiley.

Wigglesworth, G., & Foster, P. (2007). Capturing accuracy in second language performance: Exploring global and local measures. Paper presented at the 29th Annual Language Testing Research Colloquium. Universitat de Barcelona (Barcelona, Spain).

Williamsen, V. G. (1968). A pilot program in teaching Spanish: An intensive approach. *The Modern Language Journal, 52*, 73-78.

Wilson, J. M. P. (1965). Time, a constant factor in second language learning. *Hispania, 48,* 311-313.

Wilson, W. P. (1976). Developmental changes in the lag effect: An encoding hypothesis for repeated word recall. *Journal of Experimental Child Psychology, 22,* 113-122.

Winston, A., & Boots, J. (1987). A modification of the Darmouth intensive language model. *ADFL Bulletin, 19*, 18-21.

Wlodkowski, R. J. (2003). Accelerated learning in colleges and universities. *New Directions for Adult and Continuing Education, 97*, 5-15.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second Langauge Development in Writing: Measures of Fluency, Accuracy and Complexity. Technical Report 17.* Manoa, Hawai'i: University of Hawai'i Press.

Yau, M. (1991). The role of language factors in second language writing. In L. Malave & G. Duquette (Eds.), *Language, Culture and Cognition: A Collection of Studies in First and Second language Acquisition* (pp. 266-283). Clevedon, England: Multilingual Matters.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity, and accuracy in L2 monologic production. *Applied Linguistics, 24*, 1-27.

Zeydel, E. H. (1946). Introducing the Symposium. *The German Quarterly, 19*, 1-5.

APPENDICES

Appendix A: Level 3 Test

Name: _____ Date: _____
Teacher: _____ Group: _____

PART 1: Listening comprehension (Total _____ / 6 marks)

There are three sections in this part of the exam. You will hear each recording twice.

- *Look at the notes about a competition.*
- *Some information is missing.*
- *You will hear a woman talking about the competition..*
- *For each question, fill in the missing information in the number space. Each correct answer is worth one mark.*

This Month's Competiton

Prize: a computer + a (7) _____ printer

Write a story:

Length: less than (8) _____ words

Subject: a short (9) _____ story which takes place in (10) _____.

Write your name, address, telephone number and (11) _____ at the end.

Story must arrive on or before (12) _____

PART 2: Sentence conversion (5 marks)

Complete each sentence in such a way that it means the same as the sentence before it. Write your answers on the answer sheet provided.

1. They shouldn't allow them to smoke at work.

   Smoking .....

2. He tried to lift the table but he was too weak.

   He wasn’t .....

3. The weather was foggy so they couldn't go skiing.

   Because of .....

4. Shall I open the window for you?

   Would .....?

5. I can’t pay my rent today because I don’t have enough money.

   If I .....

PART 3: Cloze (5 marks)

Read the following text. Complete each gap with ONE suitable word. Contractions (don't; can't) are considered one word. Write your answers on the answer sheet provided.

A long journey

My journey to Hong Kong in the summer of 1993 was the most tiring trip I've (1) ..... been on in my life! (2) ..... all began with a flight from Barcelona to Madrid and then another (3) ..... to Paris, (4) ..... the plane stopped for two hours while people came on board (5) ..... clean it. This was the worst part of the journey because we couldn't get off the plane; we just (6) ..... to sit there and wait. The time passed very slowly.

The next stage was from Paris to Singapore, which was about thirteen hours, if I remember rightly. The service was very good, the food was delicious, there (7) ..... films on video and the flight attendants were very kind and helpful. The only problem was that I couldn't sleep at all, (8) ..... when we arrived in Singapore I was really exhausted. I tried to sleep on a sofa in the airport but there was too (9) ..... noise. Three hours later I got on the next plane, which was going from Singapore to Hong Kong and we finally arrived at about five in the afternoon. I (10) ..... been travelling for nearly thirty hours and felt like a zombie!

PART 4: Reading (10 marks)

Section A (5 marks)

*You are going to read a magazine article about some British Seaside Resorts. For questions **1-10**, choose from the sections **(A-E)**. There is only one answer for each question. The first question has been done as an example.*

*For questions **11-15,** find a word or words in the text with a similar meaning to those given. Write your answers on the answer sheet provided.*

Which resort or resorts would you recommend for someone who

likes sunbathing?	0. **D**
likes history and things from the past?	1.____ 2._____ 3.____
likes studying nature?	4.____
likes scenery?	5.____ 6._____
is elderly and likes to take things easy?	7.____
likes a busy town with lots going on?	8.____
likes amusement parks?	9.____
wants an inexpensive holiday?	10.___

Section B (5 marks)

Find <u>a word (or words)</u> in the text with a similar meaning to:

11. proudly offers (text A).....
12. successful (text B).....
13. a short holiday (text C).....
14. in good condition (text C).....
15. are included (text E).....

British Seaside Resorts

Britain is famous for its seaside resorts and, though many Britons now prefer to take their holidays in countries where the summers are sunnier, hotter and generally more reliable, Britain enjoys a yearly influx of tourists from those very countries. Britain's resorts are clearly fighting for their share of trade, and some have developed excellent weather-proof indoor attractions. Here is a selection of 5 places which are well worth a visit.A)

A) Blackpool

Blackpool is famous for its Golden Mile - a huge stretch of beautiful sandy beach. One of Britain's leading resorts, it boasts two piers and 10 miles of amazing illuminations - glittering lighted tableaux that turn the seafront into a wonderland in the autumn months. For most of the year, Blackpool's Pleasure Beach offers plucky youngsters the chance to take some terrifying rides. (Older folk can try them too if they so desire. A woman of 100 recently tried the new Big Dipper. She said she enjoyed it - though she didn't want another turn!)

B) Great Yarmouth

A little smaller than Blackpool, the town has an even wider range of things to offer. This springs from the fact that the town has other dimensions besides its tourist industry. It is a thriving industrial centre and a busy port, servicing the quest for gas and oil beneath the North Sea. It also has historical attractions and, though recent years have seen a catastrophic decline in its fishing industry, seafood can still be bought in one of England's largest market places. Compared with the ones at Blackpool, Yarmouth's swimming pool and illuminations are very modest. However, there is plenty to compensate for this.

C) Morecambe

Once a thriving resort, Morecambe now has a slightly old-fashioned, rundown air. This has its own charm and, for those desiring a relaxing break, it is hard to think of anywhere better. For its modest size, the town has most attractive shopping and eating facilities and some unspoilt surroundings. The Cumbrian coast can be seen from the promenade, providing a scenic backdrop to the happy sight of children playing (or donkey riding) on the beach.

Families and the elderly can enjoy a choice of traditional English boarding houses at prices that reflect the undeserved loss of popularity which Morecambe has suffered in recent years.

D) Brighton

Often regarded as the queen of English seaside resorts, Brighton has class! Beautiful Victorian buildings recall its magnificent past as a fashionable resort of the English gentry. However, Brighton has also moved with the times. While preserving its heritage it has cultivated a huge range of up-to-the-minute attractions, though without the gaudy vulgarity one finds at Blackpool. On the warm south coast, Brighton is a definite must for sun-seekers.

E) Sheringham

On the north Norfolk coast, Sheringham has the air of a place which is gradually coming into its own. The smallest of the five resorts which are featured here, it is gradually developing as an unpretentious and very appealing centre for a wide range of holidays. For lovers of old-style railways there is the preserved line offering steam-hauled trips to nearby Holt, a lovely little town in the North Norfolk Heights. (The train can hardly manage the steep ascent out of Sheringham and heavier trains pass non-stop through the intermediate stations, since if they stopped they might have trouble starting again.) The Heights rise to only a hundred metres, but they provide a distinct environment for a range of unusual plants and birds, making Sheringham an excellent centre for wildlife enthusiasts.

Appendix B: Level 5 Test

Name: _____ Date: _____
Teacher: _____ Group: _____

PART 1. Listening

You will hear someone giving information about how to travel around Turkey. Look at sentences 7-15 and complete them by writing one or two words in the spaces. You will hear the information twice.

It is | 7 | | to take a bus than a train.

It is possible to travel | 8 | | to Istanbul and Izmir.

The Blue Train has very comfortable | 9 | | .

It's advisable to book seats during | 10 | | .

In order to find out where the buses are going to listen to the | 11 | | .

It's not a good idea to sit on the | 12 | | during the summer.

All coaches carry | 13 | | .

From time to time you will be offered | 14 | | to refresh yourself.

Don't expect to get a good night's sleep because the bus will | 15 | | .

PART 2: Sentence conversion (5 marks)

Complete each sentence so that it means the same as the sentence before it. Write your answers on the answer sheet provided.

EXAMPLE : Shakespeare wrote 'Macbeth'

'Macbeth' was written by Shakespeare

1. I asked, "Would you like me to take you to the airport?"

   I offered …….

2. I'm sorry I spoke to her so rudely!

   I wish……..

3. Somebody will meet you at the station.

   You…….

4. Although we were tired, we finished the job.

   Despite……

5. She got angry with me because I forgot her birthday.

   If I……

PART 3: Open "cloze" (5 marks)

Read the following text. Fill in the gap with ONE word only. Contractions count as one word. Write the answers on your answer sheet.

THE ARK IN THE DOCK

A very unusual legal battle has been taking place in Australia. Allen Roberts claims to _____ (1) found remains of Noah's Ark at Akyayla in eastern Turkey, _____ (2) 20 kilometres southeast of Mount Ararat, _____ (3) is where the Ark is traditionally said to have come to rest. _____ (4) to Mr Roberts, not only are there remains of petrified wood, but also iron rivets and stone anchors. In _____ (5) to raise more money for his project, he has been giving lectures all over Australia and selling audio and video tapes of them.

Scientists, led by Professor Plimer of the University of Melbourne, have accused Roberts of manipulating the evidence; they insist that what Roberts has found _____ (6) be remains of Noah's Ark; they think the remains are part of a 100-million-year old geological formation that had _____ (7) to do with Noah and the Ark.

Plimer and his colleagues want to prevent Roberts from selling tapes of his lectures in order to raise funds. As a result, Roberts has sued Plimer for defamation and, surprisingly, Plimer has lost his case. The judge said that if Roberts had been trying to make a personal profit, then it _____ (8) be a case of deceiving the public. He also said that a court of law was not the proper place to judge the validity of competing ideas. As a _____ (9), Professor Plimer has had to sell his own home to cover his legal costs. Outside the court he said, "What I can't understand is how the judge found against us. Now I wish I _____ (10) got involved in the case."

Appendix C: Background Questionnaire (Students)

CUESTIONARIO

Apellidos y nombre: _____
E-mail: _____
Edad: _____ Sexo: Mujer Hombre
Estudios: _____
Situación laboral: Estudio Trabajo Estudio y trabajo Otras: _____
Aparte de EIM, ¿estás recibiendo otras clases de inglés? Sí No

1. ¿Cuánto tiempo llevas aprendiendo ingles?
a) 1 año o menos b) 1-5 años c) 5-10 años d) más de 10 años

2. El tiempo que has aprendido inglés ha sido
a) más o menos continuo b) interrumpido por diversas circunstancias

3. ¿A qué edad empezaste a aprender inglés? _____

4. ¿Cuándo fue la última vez que hiciste un curso de inglés? ¿Dónde fue? Por favor, especifica inicio y final (ex: octubre 2003-mayo 2004)
_____
_____
_____
_____

5. ¿Por qué estás aprendiendo inglés?
a) porque es/será necesario para trabajo b) por placer c) otros: __

6. ¿Te gusta estudiar inglés? (1=muy poco – 5= muchísimo)
1 2 3 4 5

7. ¿Por qué?
_____
_____
_____

8. Piensas que en tu clase este año se pasa más tiempo haciendo (elige 1 o varias opciones):
a) gramática/vocabulario b) conversación c) lectura d) listening
d) escritura e) otros: _____

9. Señala cuáles de estas actividades practicas en inglés y cuántas horas dedicas **por semana**.
a) deberes de clase _____
b) películas _____
c) televisión/radio _____
d) canciones _____
e) lectura (libro, periódico, internet, etc.) _____
f) escritura (no para clase) _____

g) conversación _____
h) interacciones con nativos _____
i) otros _____

10. Aproximadamente, ¿a qué porcentaje de clases has asistido?
a) 80-100%b) 50-80% c) 50% o menos

11. Consideras que en este curso, en general, has aprendido:
a) poco b) bastante c) mucho d) muchísimo

12. ¿Has estado alguna vez en algún país en el que hablaras inglés? Sí No

13. Si contestas "Sí": ¿Dónde? _____
¿Cuánto tiempo? _____

14. En la siguiente Table, evalúa tu competencia lingüística en las siguientes lenguas que conozcas: 0) Pobre 1) Buena 2) Muy buena 3) Nativa

Lengua	Listening	Speaking	Reading	Writing	Años de estudio
Castellano					
Catalán					
Inglés					
Francés					
Otra:					

15. ¿Has hecho alguna vez o estás haciendo un curso intensivo o semi-intensivo de inglés (al menos 8 horas de inglés a la semana)?
a) Sí b) No c) Cuántos en tu vida: intensivos ___ semi-intensivos ___
d) Cuántas horas a la semana: _____

Si contestas "no", ve directamente a COMENTARIOS, en la parte de abajo.
Si contestas **"SI"** aquí tienes unas preguntas sobre el/los curso/s intensivo/s o semi-intensivo/s (piensa en cómo tener más horas de inglés a la semana afecta tu aprendizaje):

	Intensivo 1	Intensivo 2 (si has hecho más de uno)	Semi-intensivo 1	Semi-intensivo 2 (si has hecho más de uno)
Cuándo				
Dónde				
Te gustó más (+), igual (=) o menos (–) que cursos no intensivos				
Aprendiste +, =, – que cursos no intensivos				
Fue +, =, – duro que cursos no intensivos				

16. ¿Cómo evaluarías tu aprendizaje en relación al tipo de programa de inglés? Por favor, especifica si has aprendido más (+), igual (=) o menos (-) en las siguientes áreas:

	Intensivo +, =, – que no intensivo	**Semi-intensivo** +, =, – que no intensivo	**Intensivo** +, =, – semi-intensivo
Listening			
Fluidez			
Pronunciación			
Entonación			
Lectura			
Escritura			
Gramática			
Vocabulario			

COMENTARIOS:

Gracias por tu colaboración. Si quieres saber tanto los resultados de tus tests y el progreso realizado desde el principio de curso hasta ahora, como los resultados generales de mi proyecto de investigación, por favor, deja tu **e-mail** y me pondré en contacto contigo.

a) Sí b) No

Appendix D: Questionnaire for Teachers

Questionnaire for Teachers

Name: _____
Age: _____
Mother tongue: _____
Number of years teaching English: _____

1. How many intensive/semi-intensive courses have you taught?

   a) 0 b) 1-5 c) more than 5

2. Please, decide if, in your opinion, students in **intensive/semi-intensive** courses learn more than (+), less than (-), or the same as (=) students in regular courses with respect to the following areas:

	INTENSIVE	SEMI-INTENSIVE
Listening		
Speaking		
Reading		
Writing		
Grammar		
Vocabulary		

What do you think about RETENTION? Do you think that retention of what the students learn is different according to whether the program they followed was intensive, semi-intensive or regular?

_____
_____
_____
_____

Other comments:

_____
_____
_____

3. You think that students in intensive courses are more (+), less (-) or as (=) students in regular courses with respect to the following characteristics:

Do

	INTENSIVE	SEMI-INTENSIVE
Motivated		
Participative		
Talented for learning languages		
Mature		
Hard-working		
Other(s)		

you think that students need special characteristics/qualities to succeed in intensive or semi-

intensive courses that are not so necessary in the case of students in regular courses?

_____
_____
_____

4. Please, order the following practice activities according to how much time you devote to them in class for intensive and regular groups, with 1 being the least commonly practiced one and 5 the most practiced (if you want you can write in parenthesis the approximate percentage of time that each type of activity is practiced in class). If some activities are given equal weight write the same number next to them.

	INTENSIVE	SEMI-INT.	REGULAR
Listening (tapes, videos, etc.)			
Grammar			
Vocabulary			
Reading			
Writing			
Speaking			

5. Please, write the number of compositions A WEEK you ask your students to write in:
   Intensive courses _____
   Semi-intensive courses _____
   Regular courses _____

6. Please, circle the group that gets THE MOST… (if same amount, you can circle several)

Homework	INTENSIVE	SEMI-INT.	REGULAR
PRACTICE in writing	INTENSIVE	SEMI-INT.	REGULAR
PRACTICE in reading	INTENSIVE	SEMI-INT.	REGULAR
PRACTICE in speaking	INTENSIVE	SEMI-INT.	REGULAR
PRACTICE in listening	INTENSIVE	SEMI-INT.	REGULAR
PRACTICE in grammar	INTENSIVE	SEMI-INT.	REGULAR
PRACTICE in vocabulary	INTENSIVE	SEMI-INT.	REGULAR

7. Please, add any comment with respect to your opinion, feelings or impressions about intensive/semi-intensive English courses and how they are different from the regular courses you have taught.

Thanks a lot for your collaboration

Appendix E: Scores Listening Excercise Level 5

	1 point	0.5 points	0 points
Blank 1	Cheaper Faster Quicker More convenient	Convenient Better	Costless
Blank 2	Overnight During the night At night In night service During the day and night		
Blank 3	Reclining seats	Reclined seats Reclaining seats Seats	Sits
Blank 4	Public holidays Turkish public holidays Turkish holidays	Popular holidays Holidays	
Blank 5	Men shouting	Shouts Men	
Blank 6	Sunny side Sun side	Shining side	
Blank 7	Bottled water Botteled water	Bottle water Cold water Water Water bottles, etc.	Refrigerator container Refrigery container
Blank 8	Lemon cologne Lemon colone	Lemon Cologne Lemon colon	
Blank 9	Stop frequently Often stop	Stop	Move a lot

Appendix F: Example of Transcription

@Begin

@Languages: en

@Participants: MAR, Maria_Pujol RAQ, Raquel Investigator

@ID: 3

@Coder: Raquel

@Tape Location: PreRA1S2_289

*PAQ: I don't have an only best friend because I think that [err] is
difficult that one person have [err] all the qualities that
I want [T] [CL] [CL] [CL] [CL] [CL] .

*PAQ: If I have to choose I think that my best friend is my sister [EFTU]
[CL] [CL] [CL] .

*PAQ: She has [err] 31 years [T] [CL] .

*PAQ: and we are very similar [EFTU] [CL] .

*PAQ: She always listen [err] [T] [CL] .

*PAQ: and help [err] me in all that she can [T] [CL] [CL] .

*PAQ: She always tell [err] me my mistakes [T] [CL] .

*PAQ: and never lie [err] [err] me [T] [CL] .

*PAQ: We keep in touch although I'm studying in Barcelona [EFTU]
[CL] [CL] .

*PAQ: She phoned me for simply talk [err] [T] [CL] [CL] .

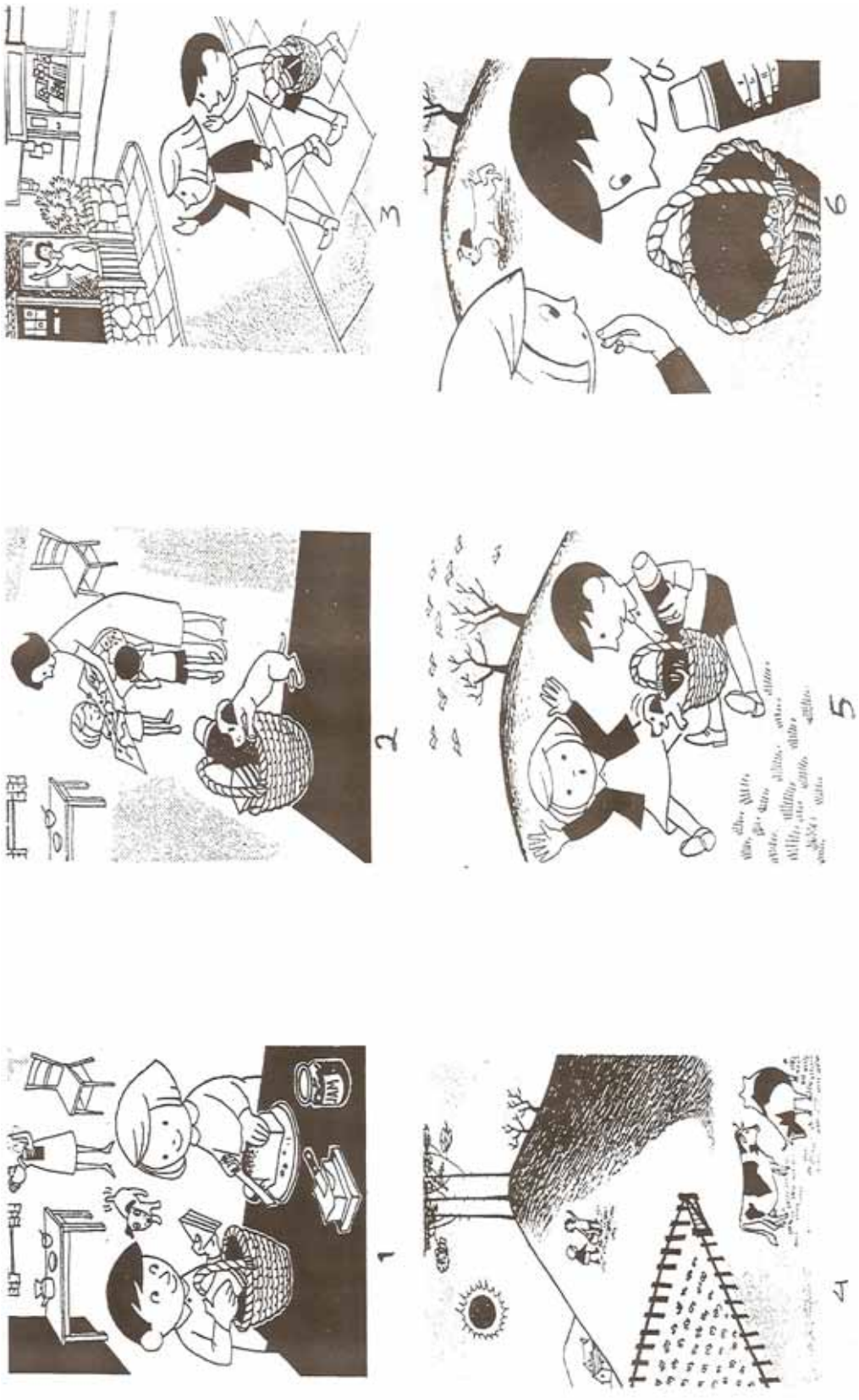
*PAQ: and this is very important [EFTU] [CL] .

*PAQ: In December [err] will travel together to England [T] [CL] .

*PAQ: and I hope that we enjoy [err] very much [T] [CL] [CL] .

@End

Appendix G: Oral Narrative (The Dog's Story)



Appendix H: Non-Parametric Tests Level 3 Written Test

As the following tables illustrate, the results of the Levene's Tests indicated that the assumption of equality of variance was violated in the case of the listening, the reading B exercise, and the reading total in the posttest.

Levene's Test of Equality of Error Variances(a)

	F	Df1	df2	Sig.
Prelist	.509	2	128	.602
Postlist	5.800	2	128	.004

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreSentConv	.936	2	128	.395
PostSentConv	1.402	2	128	.250

Levene's Test of Equality of Error Variances(a)

	F	df1	Df2	Sig.
PreCloze	.273	2	128	.761
PostCloze	1.167	2	128	.314

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreReadA	.138	2	128	.871
PostReadA	.126	2	128	.881

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreReadB	.041	2	128	.959
PostReadB	12.727	2	128	.000

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreReadTot	.346	2	128	.708
PostReadTot	7.120	2	128	.001

Levene's Test of Equality of Error Variances(a)

	F	Df1	df2	Sig.
PreTot	.202	2	128	.817
PostTot	1.455	2	128	.237

The samples for almost all the groups for all the different exercises included in the written test (except for the total scores) failed the Kolmogorov-Smirnov test of normality.

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	Df	Sig.
Prelist	3Ext	.113	49	.155
	3Semi	.158	44	.008
	3Int	.145	38	.042
PostList	3Ext	.203	49	.000
	3Semi	.244	44	.000
	3Int	.190	38	.001

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreSentConv	3Ext	.242	49	.000
	3Semi	.225	44	.000
	3Int	.200	38	.001
PostSentConv	3Ext	.185	49	.000
	3Semi	.144	44	.022
	3Int	.151	38	.028

Tests of Normality

	Program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreCloze	3Ext	.124	49	.055
	3Semi	.153	44	.011
	3Int	.168	38	.008
PostCloze	3Ext	.154	49	.005
	3Semi	.159	44	.007
	3Int	.193	38	.001

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreReadA	3Ext	.145	49	.012
	3Semi	.189	44	.000
	3Int	.154	38	.023
PostReadA	3Ext	.155	49	.005
	3Semi	.219	44	.000
	3Int	.210	38	.000

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreReadB	3Ext	.268	49	.000
	3Semi	.190	44	.000
	3Int	.243	38	.000
PostReadB	3Ext	.253	49	.000
	3Semi	.210	44	.000
	3Int	.186	38	.002

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreReadTot	3Ext	.167	49	.002
	3Semi	.092	44	.200(*)
	3Int	.165	38	.010
PostReadTot	3Ext	.125	49	.052
	3Semi	.101	44	.200(*)
	3Int	.095	38	.200(*)

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreTot	3Ext	.071	49	.200(*)
	3Semi	.092	44	.200(*)
	3Int	.089	38	.200(*)
PostTot	3Ext	.074	49	.200(*)
	3Semi	.104	44	.200(*)
	3Int	.079	38	.200(*)

Since most samples failed the Kolmogorov-Smirnov test, other normality tests were performed, as well as non-parametric tests. Another way to test for normality is by dividing the skewness and the kurtosis by their standard errors. If the result of such calculation is below 2 or -2, such sample can be said to be normal. The skewness and the kurtosis for the different groups in each exercise can be seen in Table 6.1. The standard error of the skewness for the extensive group is .340, for the semi-intensive .357; for the intensive .383. The standard error for the kurtosis for the extensive, semi-intensive and intensive is .668, .702 and .750 respectively. The results of the division are presented in Table 8.1.

Table 8.1: Another Normality Test

		Extensive		Semi-intensive		Intensive	
		Sk/err	Kur/err	Sk/err	Kur/err	Sk/err	Kur/err
PRETEST	**List /6**	-1.48	.029	**-3.29**	**3.21**	-1.06	-.333
	SC /5	**3.61**	.983	**3.39**	.668	**5.77**	**10.25**
	Cloze /5	.226	-.820	-1.21	-.061	-1.17	-.800
	Read1 /5	-.170	-.393	-.795	.480	-.407	.190
	Read2 /5	1.21	-1.37	.266	-1.28	1.15	-1.23
	ReadT /10	.856	-.673	.081	-.112	.211	-.938
	TOTAL /26	.002	-1.26	.201	-.962	-.365	.677
POSTTEST	**List /6**	-1.60	-.985	**-3.81**	**2.36**	**-4.02**	**3.64**
	SC /5	**2.35**	-.552	1.10	**-3.21**	1.99	-.396
	Cloze /5	-1.22	-.845	**-2.24**	.242	-1.55	-.694
	Read1 /5	-.091	-.093	-1.73	.417	-.770	-.350
	Read2 /5	**2.43**	-.014	.826	-1.82	-.216	-.154
	ReadT /10	1.38	.941	.263	-1.65	-.553	.180
	TOTAL /26	.050	-1.30	.308	-1.66	-.242	.549

It can be observed in Table 8.1 that this normality test still fails for some groups in some of the exercises (listening, sentence conversion, and marginally cloze and reading B), that is why non-parametric tests were performed for those measures so as to analyze whether similar results were obtained as in the parametric tests regarding statistical significance.

In order to examine whether there were differences between the three programs in the pretest and in the posttest in the listening, sentence conversion, cloze and reading B, Kruskal-Wallis tests were performed with the three program types as independent variables. Subsequently, Mann-Whitney *U* tests were executed to compare between two groups when statistically significant results appeared in the former test.

The results of the Kruskal-Wallis test revealed that there were no significant differences in the pretest among groups in the listening (χ^2 3.62, *df* 2, p = .164), sentence conversion (χ^2 .950, *df* 2, p = .622), cloze (χ^2 2.25, *df* 2, p = .325) or reading B (χ^2 6.05, *df* 2, p = .187), which confirm the findings reported in Table 6.2. On the other hand, significant

differences were found in the posttest in the listening (χ^2 6.63, *df* 2, *p* = .036), sentence conversion (χ^2 6.01, *df* 2, *p* = .049), reading B (χ^2 6.05, *df* 2, *p* = .049), but not in the cloze (χ^2 3.63, *df* 2, *p* = .163). These results are practically the same as those presented in Table 6.2, with the exception that with the two-way mixed ANOVA no significant difference was found in the reading B in the posttest; however, the *p* value is .089 on the ANOVA and .049 on the Kruskal Wallis.

Mann-Whitney *U*-tests were performed so as to determine more specifically where the statistical significance lay. The results from this test indicate that the extensive and semi-intensive group were significantly different in the posttest only in the listening (*U* 783, *Z* -2.25, *p* = .019). The differences in the sentence conversion, cloze and reading B were not statistically significant (*U* 874, *Z* -1.60, *p* = .110; *U* 887.5, *Z* -1.485, *p* = .138; *U* 1006, *Z* -.573, *p* = .567 respectively).

Conversely, the results of the Mann-Whitney *U*-test revealed that the differences between the performance of the students in the extensive and intensive program were marginally significant in the posttest in all the exercises under consideration, except for the cloze (listening: *U* 710, *Z* -1.94, *p* = .052; sentence conversion: *U* 652.5, Z -2.413, *p* = .016; cloze *U* 733, *Z* -1.717, *p* = .086; reading B: *U* 614.5, *Z* -2.81, *p* = .005).

The results of the students in the semi-intensive and the intensive program were not significantly different in the posttest in the listening (*U* 785, *Z* -.495, *p* = .621), sentence conversion (*U* 758.5, *Z* -.727, *p* = .467), cloze (*U* 795.5, *Z* -.383, *p* = .702), nor in the reading B (*U* 711.5, *Z* -1.18, *p* = .238).

When analyzing the *time* variable, the results of the Wilcoxon Signed Rank Test for each group with the scores on the pre and posttest for each of the measures as dependent variable revealed that the performance in the posttest in the listening exercise was significantly different for students in extensive (*Z* -2.31, *p* =.021), semi-intensive (*Z* -3.00, *p*

=.003), and intensive programs (*Z* -4.27, *p* =.000). The results were also significantly different in the sentence conversion: extensive (*Z* -2.35, *p* =.019), semi-intensive (*Z* -3.88, *p* =.000), and intensive programs (*Z* -4.23, *p* =.000). Similarly, all the students obtained significantly higher scores in the posttest in the cloze activity: extensive (*Z* -3.29, *p* =.001), semi-intensive (*Z* -4.42, *p* =.000) and intensive programs (*Z* -3.63, *p* =.000), and in the reading B: extensive (*Z* -2.78, *p* = .005), semi-intensive (*Z* -2.33, *p* =.020), and intensive (*Z* -4.18, *p* = .000) programs. These results are practically the same as the ones obtained using the two-way mixed ANOVA presented in the column 'Pairwise comparisons' in Table 6.2.

Appendix I: Normality and Levene's Test Level 3 Writing

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreW/T	3Ext	.074	40	.200(*)
	3Semi	.121	31	.200(*)
	3Int	.082	32	.200(*)
PostW/T	3Ext	.133	40	.072
	3Semi	.076	31	.200(*)
	3Int	.132	32	.165

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreC/T	3Ext	.101	40	.200(*)
	3Semi	.105	31	.200(*)
	3Int	.119	32	.200(*)
PostC/T	3Ext	.133	40	.071
	3Semi	.074	31	.200(*)
	3Int	.118	32	.200(*)

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreGuiraud	3Ext	.130	40	.087
	3Semi	.080	31	.200(*)
	3Int	.103	32	.200(*)
PostGuiraud	3Ext	.078	40	.200(*)
	3Semi	.068	31	.200(*)
	3Int	.101	32	.200(*)

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreEFTU/T	3Ext	.129	40	.092
	3Semi	.102	31	.200(*)
	3Int	.173	32	.016
PostEFTU/T	3Ext	.067	40	.200(*)
	3Semi	.109	31	.200(*)
	3Int	.118	32	.200(*)

Tests of Normality[53]

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreErr/T	3Ext	.140	40	.048
	3Semi	.183	31	.009
	3Int	.190	32	.005
PostErr/T	3Ext	.083	40	.200(*)
	3Semi	.143	31	.105
	3Int	.136	32	.141

[53] The errors per T-unit was the only measure where there was a slight abnormal distribution in the pretest once the outliers were removed, because of the distribution in the semi-intensive group. The EFTU/T was slightly abnormal according to the Kolmogorov-Smirnov test, but not according to the result of dividing the skewness and kurtosis by their standard errors. The results of the non-parametric tests for the Err/T measure confirmed the outcome of the two-way mixed ANOVA. In the pretest the Kruskal-Wallis Test evidenced no significant differences across programs (χ^2 3.37, *df* 2, p = .185); nevertheless, the results in the posttest were significantly different (χ^2 13.69, *df* 2, p = .001). The Wilcoxon Signed Rank tests revealed that there were no significant differences in the performance of the students in semi-intensive (Z -.274, p = -.784) and intensive (Z .197, p = .197) programs; however, the students in the extensive program showed significantly less accurate results in the posttest (Z -2.812, p = .005). All of these results are in agreement with the ANOVA.

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreW/T	1.909	2	100	.154
PostW/T	3.725	2	100	.028

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreC/T	4.223	2	100	.017
PostC/T	2.060	2	100	.133

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreGuiraud	.555	2	100	.576
PostGuiraud	.277	2	100	.759

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreEFTU/T	1.084	2	100	.342
PostEFTU/T	2.095	2	100	.128

Levene's Test of Equality of Error Variances(a)

	F	Df1	df2	Sig.
PreErr/T	2.115	2	100	.126
PostErr/T	.911	2	100	.405

Appendix J: Normality Test Level 3 Speaking

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
Presyll/min	3Ext	.108	21	.200(*)
	3Semi	.169	11	.200(*)
	3Int	.129	20	.200(*)
Postsyll/min	3Ext	.093	21	.200(*)
	3Semi	.099	11	.200(*)
	3Int	.128	20	.200(*)
PreC/T	3Ext	.259	21	.001
	3Semi	.120	11	.200(*)
	3Int	.164	20	.164
PostC/T	3Ext	.208	21	.018
	3Semi	.223	11	.133
	3Int	.165	20	.155
PreGiraud	3Ext	.102	21	.200(*)
	3Semi	.156	11	.200(*)
	3Int	.139	20	.200(*)
PostGiraud	3Ext	.125	21	.200(*)
	3Semi	.112	11	.200(*)
	3Int	.141	20	.200(*)
PreEFTU/T	3Ext	.184	21	.061
	3Semi	.107	11	.200(*)
	3Int	.156	20	.200(*)
PostEFTU/T	3Ext	.176	21	.087
	3Semi	.187	11	.200(*)
	3Int	.133	20	.200(*)
PreErr/T	3Ext	.134	21	.200(*)
	3Semi	.221	11	.141
	3Int	.170	20	.132
PostErr/T	3Ext	.087	21	.200(*)
	3Semi	.195	11	.200(*)
	3Int	.216	20	.015

Appendix K: Normality and Levene's Level 5 Written Test

Tests of Normality

	Program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreList	5Ext	.160	37	.018
	5Int	.121	31	.200(*)
PostList	5Ext	.105	37	.200(*)
	5Int	.119	31	.200(*)

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreSentConv	5Ext	.186	34	.004
	5Int	.160	31	.043
PostSentConv	5Ext	.147	34	.061
	5Int	.148	31	.081

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreCloze	5Ext	.162	37	.015
	5Int	.140	31	.128
PostCloze	5Ext	.182	37	.003
	5Int	.180	31	.012

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreTot	5Ext	.075	34	.200(*)
	5Int	.081	31	.200(*)
PostTot	5Ext	.097	34	.200(*)
	5Int	.108	31	.200(*)

Since for some measures the Kolmogorov-Smirnov test of normality failed, the skewness and kurtosis were divided by their standard errors. The results of such

computation are represented in the following table, and it can be seen that such results are within the range of +2 and -2, which is why the distribution of all the measures can be said to be normal.

Another Normality Test Written Test Level 5

		Extensive		Intensive	
		Sk/err	Kur/err	Sk/err	Kur/err
PRETEST	**List /6**	.456	-.840	1.23	-.746
	SC /5	.806	-.585	.862	-.097
	Cloze /5	-.757	0	.168	-.713
	TOTAL /26	.163	-.322	-.201	-1.00
POSTTEST	**List /6**	-.615	.466	-.427	-.650
	SC /5	.220	-1.05	-1.14	-.041
	Cloze /5	-.072	-1.15	-1.12	.046
	TOTAL /26	-.129	-.491	-.950	-.605

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreList	1.157	1	66	.286
PostList	.488	1	66	.487

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreSentConv	.000	1	63	.999
PostSentConv	.933	1	63	.338

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreCloze	.013	1	66	.911
PostCloze	.466	1	66	.497

Levene's Test of Equality of Error Variances(a)

	F	Df1	df2	Sig.
PreTot	.125	1	63	.725
PostTot	.135	1	63	.715

Appendix L: Normality and Levene's Tests Level 5 Writing

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreW/T	5Ext	.089	32	.200(*)
	5Int	.163	31	.035
PostW/T	5Ext	.108	32	.200(*)
	5Int	.082	31	.200(*)

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreC/T	5Ext	.108	32	.200(*)
	5Int	.135	31	.157
PostC/T	5Ext	.137	32	.134
	5Int	.107	31	.200(*)

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreGuiraud	5Ext	.199	32	.002
	5Int	.091	31	.200(*)
PostGuiraud	5Ext	.084	32	.200(*)
	5Int	.106	31	.200(*)

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreEFTU/T	5Ext	.089	32	.200(*)
	5Int	.080	31	.200(*)
PostEFTU/T	5Ext	.102	32	.200(*)
	5Int	.109	31	.200(*)

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreErr/T	5Ext	.085	32	.200(*)
	5Int	.089	31	.200(*)
PostErr/T	5Ext	.148	32	.072
	5Int	.143	31	.105

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreW/T	3.066	1	61	.085
PostW/T	.130	1	61	.720

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreC/T	1.984	1	61	.164
PostC/T	.174	1	61	.678

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreGuiraud	2.035	1	61	.159
PostGuiraud	.045	1	61	.833

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreEFTU/T	.685	1	61	.411
PostEFTU/T	.066	1	61	.798

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
PreErr/T	.919	1	61	.342
PostErr/T	.345	1	61	.559

Appendix M: Normality Test Level 5 Speaking

Tests of Normality

	program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
Presyll/min	5	.128	23	.200(*)
	51	.131	22	.200(*)
Postsyll/min	5	.122	23	.200(*)
	51	.248	22	.001
PreGiraud	5	.094	23	.200(*)
	51	.118	22	.200(*)
PostGiraud	5	.143	23	.200(*)
	51	.120	22	.200(*)
PreEFTU/T	5	.126	23	.200(*)
	51	.110	22	.200(*)
PreC/T	5	.161	23	.128
	51	.182	22	.057
PostC/T	5	.084	23	.200(*)
	51	.195	22	.029
PostEFTU/T	5	.232	23	.002
	51	.113	22	.200(*)
PreErr/T	5	.123	23	.200(*)
	51	.153	22	.195
PostERR/T	5	.190	23	.030
	51	.118	22	.200(*)

Appendix N: Normality and Levene's Tests L3 vs. L5 Writing

Tests of Normality

	level	Kolmogorov-Smirnov(a)		
		Statistic	Df	Sig.
PreW/T	3	.064	103	.200(*)
	5	.119	63	.028
PostW/T	3	.108	103	.005
	5	.062	63	.200(*)

Tests of Normality

	level	Kolmogorov-Smirnov(a)		
		Statistic	Df	Sig.
PreC/T	3	.077	103	.151
	5	.128	63	.013
PostC/T	3	.109	103	.004
	5	.107	63	.072

Tests of Normality

	level	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreGuiraud	3	.052	103	.200(*)
	5	.141	63	.003
PostGuiraud	3	.049	103	.200(*)
	5	.047	63	.200(*)

Tests of Normality

	level	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreEFTU/T	3	.073	103	.200(*)
	5	.078	63	.200(*)
PostEFTU/T	3	.076	103	.161
	5	.099	63	.200(*)

Tests of Normality

	Level	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreErr/T	3	.149	103	.000
	5	.057	63	.200(*)
PostErr/T	3	.075	103	.169
	5	.129	63	.011

Levene's Test of Equality of Error Variances(a)

	F	df	Sig.
PreW/T	54.627	164	.000
PostW/T	16.726	164	.000

Levene's Test of Equality of Error Variances(a)

	F	df	Sig.
PreC/T	34.205	164	.000
PostC/T	6.346	164	.013

Levene's Test of Equality of Error Variances(a)

	F	df	Sig.
PreGuiraud	.772	164	.381
PostGuiraud	2.562	164	.111

Levene's Test of Equality of Error Variances(a)

	F	df	Sig.
PreEFTU/T	.016	164	.898
PostEFTU/T	1.692	164	.195

Levene's Test of Equality of Error Variances(a)

	F	df	Sig.
PreErr/T	.363	164	.547
PostErr/T	1.155	164	.284

Appendix O: Normality and Levene's Tests L3 vs. L5 Speaking

Tests of Normality

	level	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreSyll/min	3	.059	53	.200(*)
	5	.078	45	.200(*)
PostSyll/min	3	.084	53	.200(*)
	5	.151	45	.012

Tests of Normality

	level	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreC/T	3	.194	53	.000
	5	.190	45	.000
PostC/T	3	.112	53	.095
	5	.118	45	.125

Tests of Normality

	level	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreGuiraud	3	.080	53	.200(*)
	5	.058	45	.200(*)
PostGuiraud	3	.083	53	.200(*)
	5	.103	45	.200(*)

Tests of Normality

	level	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreEFTU/T	3	.125	53	.039
	5	.098	45	.200(*)
PostEFTU/T	3	.115	53	.077
	5	.113	45	.189

Tests of Normality

	level	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreErr/T	3	.139	53	.012
	5	.127	45	.067
PostErr/T	3	.113	53	.086
	5	.109	45	.200(*)

Levene's Test of Equality of Error Variances(a)

	F	df	Sig.
Presyll/minB	1.670	96	.199
Postsyll/minB	5.045	96	.027

Levene's Test of Equality of Error Variances(a)

	F	df	Sig.
PreC/T	1.995	96	.161
PostC/T	.067	96	.796

Levene's Test of Equality of Error Variances(a)

	F	df	Sig.
PreGuiraud	.195	96	.660
PostGuiraud	1.055	96	.307

Levene's Test of Equality of Error Variances(a)

	F	df	Sig.
PreEFTU/T	1.682	96	.198
PostEFTU/T	.123	96	.727

Levene's Test of Equality of Error Variances(a)

	F	df	Sig.
PreErr/T	4.966	96	.028
PostErr/T	2.391	96	.125

Appendix P: Normality Test L5 vs. NESs Writing

Tests of Normality

	Program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
PreW/T	5Ext	.089	32	.200(*)
	NESs	.274	13	.009
	5Int	.163	31	.035
PostW/T	5Ext	.108	32	.200(*)
	NESs	.274	13	.009
	5Int	.082	31	.200(*)
PreC/T	5Ext	.108	32	.200(*)
	NESs	.251	13	.024
	5Int	.135	31	.157
PostC/T	5Ext	.137	32	.134
	NESs	.251	13	.024
	5Int	.107	31	.200(*)
PreGuiraud	5Ext	.199	32	.002
	NESs	.227	13	.066
	5Int	.091	31	.200(*)
PostGuiraud	5Ext	.084	32	.200(*)
	NESs	.227	13	.066
	5Int	.106	31	.200(*)
PreEFTU/T	5Ext	.089	32	.200(*)
	NESs	.307	13	.002
	5Int	.080	31	.200(*)
PostEFTU/T	5Ext	.102	32	.200(*)
	NESs	.307	13	.002
	5Int	.109	31	.200(*)

Appendix Q: Normality Test L5 vs. NESs Speaking

Tests of Normality

	Program	Kolmogorov-Smirnov(a)		
		Statistic	df	Sig.
Presyll/Min	5Ext	.128	23	.200(*)
	NESs	.147	12	.200(*)
	5Int	.131	22	.200(*)
Postsyll/Min	5Ext	.122	23	.200(*)
	NESs	.147	12	.200(*)
	5Int	.248	22	.001
PreC/T	5Ext	.161	23	.128
	NESs	.179	12	.200(*)
	5Int	.182	22	.057
PostC/T	5Ext	.084	23	.200(*)
	NESs	.179	12	.200(*)
	5Int	.195	22	.029
PreGuiraud	5Ext	.094	23	.200(*)
	NESs	.178	12	.200(*)
	5Int	.118	22	.200(*)
PostGuiraud	5Ext	.143	23	.200(*)
	NESs	.178	12	.200(*)
	5Int	.120	22	.200(*)
PreEFTU/T	5Ext	.126	23	.200(*)
	NESs	.530	12	.000
	5Int	.110	22	.200(*)
PostEFTU/T	5Ext	.232	23	.002
	NESs	.530	12	.000
	5Int	.113	22	.200(*)
PreErr/T	5Ext	.123	23	.200(*)
	NESs	.499	12	.000
	5Int	.153	22	.195
PostErr/T	5Ext	.190	23	.030
	NESs	.499	12	.000
	5Int	.118	22	.200(*)