SECOND LANGUAGE FLUENCY AND COGNITION: THE STUDY OF SPANISH SECOND LANGUAGE DEVELOPMENT IN AN OVERSEAS IMMERSION PROGRAM AND AN AT-HOME FOREIGN LANGUAGE CLASSROOM

Lorenzo J. García-Amaya

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Spanish & Portuguese
Indiana University

Abstract

This dissertation investigates the development of oral fluency and L2-specific measures of cognitive abilities for two groups of L2 learners of Spanish (L1 English) in two different learning contexts: a 7-week overseas intensive immersion program (IM) in León, Spain and a 15-week domestic foreign language classroom in an at-home (AH) context at a large Midwestern institution in the United States. In total, 56 native speakers of English participated in the study – 27 in the IM program and 29 in the AH program. All learners performed a video-retell oral production task in addition to a detailed language contact profile and a proficiency test, and IM learners performed three cognitive tasks designed to elicit L2-specific measures of lexical access, lexical retrieval, and attention control. Data collection was longitudinal for both learner groups. Overall, the findings show significantly greater fluency gains for IM learners over AH learners, which can be attributed to the significantly greater amount of exposure of IM learners to the L2, as indicated in the language contact profile. In terms of cognitive tasks, IM learners show significantly faster Spanish lexical access over time, but significantly slower English lexical access over time. However, no significant longitudinal differences were seen for IM learners regarding lexical retrieval in Spanish (which requires articulation and morphophonological and phonetic encoding in addition to lexical access). The results have implications for models of speech production and processing and their applications to L2 acquisition; they also prove the methodological importance of collecting data in situ instead of after learners' return to their country of origin. Finally, this dissertation is designed to account for the role of context of learning in second language acquisition.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Introduction and research goals

In recent years we have seen a considerable increase in publications on the effects of the study abroad (SA) experience on Spanish second language acquisition (SLA). Most studies focus on the description of the SA setting by itself along with an analysis of the performance of groups of learners participating in them (e.g., Gunterman, 1992a,b; Isabelli, 2001; Isabelli, 2004; Llanes & Muñoz, 2009; Lord, 2006, 2009; Ryan & Lafford, 1995; Simões, 1996) or on a quantitative comparison of the linguistic performance of L2 learners participating in the SA context to other learners exposed to the at-home (AH) context (e.g., Collentine, 2004; DeKeyser, 1986, 1990, 1991, Díaz-Campos, 2004, 2006; García-Amaya 2008, 2009; Segalowitz & Freed, 2004; Stevens, 2001; Sunderman & Kroll, 2009, etc.). The latter group of investigations is certainly the most scientifically appealing for a researcher in SLA, as such studies typically provide a control group as a basis of comparison so that the direct effect of the SA experience can be assessed. Before publication of the special 2004 issue of *Studies in Second Language Acquisition* on learning contexts (edited by Barbara Freed and Norman Segalowitz), however, SA and AH were the only learning contexts that had received interest from an L2 research perspective. For the first time, this special issue included a study that analyzed different aspects of L2 learner performance during a US-based *immersion* (IM) program (Freed, Segalowitz, & Dewey, 2004).

IM learning contexts are particularly appealing to the L2 researcher for a variety of reasons. First, in this particular context, learners are exposed to the target language on a daily basis for considerably more hours than in the traditional AH context. IM contexts also normally include more hours of formal instruction than typical semesters abroad, where for example, college students attend classes a few days per week. The duration of IM programs is normally

shorter than the SA ones, the former ranging between seven and eight weeks while the latter typically consist of one or two semesters abroad. There is also a particular motivational aspect involved in IM programs, as the admission and selection processes are typically highly selective and more costly than enrolling in a college-based foreign language classroom. Most importantly, in some of these programs, learners abide by a language pledge to speak the target language only, and use of the L1 is avoided except in extreme circumstances. These special norms typically require a certain level of maturity as well as a real commitment on part of the learner to devote most of his/her communication time to processing the target language.

1.2 Background

1.2.1 Second language fluency

1.2.1.1 What is meant by 'fluency' and why is it interesting?

Fluency is a multidimensional concept that encompasses not only aspects related to the way speech is produced, but also to the way it is perceived. The amount of characteristics that might help a researcher to label someone within the range of 'highly fluent' to 'non-fluent' are extensive and include aspects from rate of speech to foreign accentedness, the specific vocabulary used, and the specific organization of the discourse among many other dimensions of speech. Thus, the word *fluency* can be very well labeled as an umbrella term for different aspects of speech. Freed (1995), for example, shows that results of an informal campus survey indicated that almost everyone who was asked about the definition of 'fluency' gave a different definition. Overall, participants defined fluent as 'smooth speech', 'lacking hesitations', 'without interruptions', and 'native speech'. Certainly, these definitions also resonate in the research that has been carried out on L2 fluency. Specifically, research on L2 fluency has tried to account for fluency development in the interlanguage of second language learners with both quantitative and

qualitative data (Lennon, 1990; Riggenbach, 1989, 1991; Wiese, 1982, 1984) and also with cross-sectional and longitudinal data (Lennon, 1990; Wiese, 1982, 1984).

Specifically, this line of research has investigated L2 fluency development in different learning contexts, that is, SA vs. AH. While the abroad stay takes place in a country where the target language is spoken natively, the AH context is typically characterized as a traditional foreign language instruction setting which takes place in the country of origin of the learner. Since fluency can only be accessed and analyzed through the observation of oral speech production, the core of L2 fluency research has focused on the measurement of temporal variables such as rate of speech, number of words spoken (Lennon, 1990; Freed, 1998; Raupach, 1980), the perturbations found in the speech chain, also known as speech errors or *disfluencies* (Bortfeld, 2001; Temple, 1992), and native judgments on L2 fluency (Derwing & Munro, 1997; Lennon, 2000; Munro & Derwing, 1994, 1995, 1999). Research on L2 fluency has also focused on the implications of task effect on fluency as well as how planning affects L2 oral production (Yuan & Ellis, 2003). Finally, other investigations have compared oral production in both L1 and L2 (García-Amaya, 2009). Valls-Ferrer (2011) also investigated the development of speech rhythm in L2 learners of English in an SA context (see also Mora & Valls-Ferrer, in press).

In short, the different linguistic capacities that underlie overall fluency are important aspects for interlanguage development. For instance, the fact that an learner's overall rate of speech increases from 2.20 syllables per second in third year Spanish to 3.50 syllables after spending a semester abroad (results from García-Amaya, 2009) is directly related not only to the acquisition of grammatical, lexical items, collocations and other specifics of the target language, but also to the *automatization* of cognitive processes that make these elements available to the L2 learners during online oral production. Quantification of the temporal aspects of L2

development is of extreme importance for SLA research in general because they are connected to overall L2 development. Learners would not be able to access and retrieve words faster, produce varied vocabulary and clause type and be more grammatically accurate, for example, in independent steps. On the contrary, all these aspects of interlanguage development take place in 'parallel' continuums, and examining overall fluency, whether analyzed through temporal variables of speech or hesitation phenomena is a reflection of this process. Measuring these steps at different points throughout the process of interlanguage development may be viewed as another method of documenting L2 interlanguage development. A great deal of interest has been seen recently in L2 research to further explore individual differences and L2 oral performance, an interest that springs from the understanding of oral fluency as an integral component of interlanguage development.

1.2.1.2 History of L2 fluency research

Investigations on fluency in oral speech production dates back to the 1960's with the pioneering pausological investigations conducted by Goldman-Eisler on L1 oral production (e.g., Goldman-Eisler, 1951, 1961, 1968, 1972; Henderson, Goldman-Eisler, & Skarbek, 1966). These studies did not focus on L2 fluency, as it is understood today, nonetheless, they were the seed for a subsequent number of investigations on L2 speech that would appear in the next two decades. Indeed, the 1980's brought advancement in this particular line of research with both theoretical and applied approaches to the subject (Dechert, 1980; Grosjean 1980a,b) to the point that even edited volumes on the topic were also published (Dechert, Möhle, & Raupach, 1984; Dechert & Raupach, 1987). It is in the 1990's, however, when the sub-field of fluency research grew with most intensity and, as Segalowitz points out in the latest the state-of-the-art of L2 fluency research (Segalowitz, 2010), when the language testing literature began to consider the

implications that contemporary L2 fluency research may have on the more applied assessment purposes (e.g., Bachman, 1990; Bachman & Palmer, 1996; McNamara, 1996) (p. 30).

A point of particular importance for L2 fluency research also comes in this decade with the publication of Lennon (1990). A classic in this subfield of SLA, Lennon's pioneering study has been an extremely influential study for fluency research since its publication. Both its approach to fluency analysis and the variables included have worked as a template for a plethora of researchers investigating second language fluency at large, including approaches to its definition, operationalization, measurements, assessments and by extension to identify and investigate which factors or indexes raters consciously or unconsciously perceive as fluency markers. According to Lennon, fluency mirrors listeners' judgments of specific aspects of L2 oral production, and by this definition, *fluency* is understood as an auditory and perceptual phenomenon. Indeed, Lennon tried to identify markers (or indicators) of fluency in a strict sense and thus, he included in his analysis a variety of fluency measures including: *unpruned* (with self-corrections, etc.) and *pruned* (with self-corrections excluded) speech rate measured in words per minute; interruptions, which included repetitions, self-corrections and filled pauses; percentage of repeated and self-corrected words; two pause measures; mean length of speech runs (number of words) between pauses; and finally, three measures pertaining to T-units (percent of T-units followed by a pause, percent of total pause time at all T-unit boundaries, and mean pause time at T-unit boundaries).

In an attempt to focus on very specific components of speech, Lennon left out other aspects of important consideration in speech such as accuracy, complexity, idiomaticity, pronunciation, and accentedness among others.[1] A plethora of recent investigations have

[1] These factors would later be shown to also have an impact on listener's perception of L2 fluency (Derwing et al., 2004), and focused on fluency as a matter of speed and efficiency of delivery (i.e., lack of hesitancies), two concepts

addressed these components of speech (Ellis, 2009; Ortega & Norris, 2008; Skehan, 1999, etc.), however Lennon's original approach to the study of L2 fluency has been followed by many researchers interested in the relationship between temporal and hesitation phenomena in the oral production of L2 learners (Freed, 1995; Freed, Segalowitz, & Dewey, 2004; García-Amaya, 2008, 2009; Mehnert, 1998; O'Brien, Segalowitz, Freed, & Collentine, 2007; Ortega, 1999; Riggenbach, 1991; Segalowitz & Freed, 2004; Temple, 1992, 2000, 2005). A good number of these studies have taken into the account the SA variable since the inception of the subfield. Relevant studies on this context have appeared since Carroll (1967).

Focusing more on the specific of Lennon (1990), four learners of L2 English (L1 German) were the participants of the study and were recorded before and after a six-month SA experience in England. The data were transcribed and analyzed for the aforementioned fluency indexes. In order to assess if L2 speech production had changed over this period of time, and also in an effort to stress the auditory reality in which fluency is perceived and assessed by listeners, Lennon also included native speaker judges (a practice that since then has been replicated in many fluency research investigations). When reporting quantitative results Lennon indicated that while learners' use of pausing time between Times 1 and 2 had decreased significantly, their rate of speech (measured in words per minute) had become faster. He also claimed that position, length, and frequency of pauses were at the base of learners' improvement of speech rate. He pointed out that the use of repetitions and filled pauses had more in common with one another and could be better indicators of fluency, than, for instance, the use of self-correction. Lennon did not view self-correction as indicative of fluency improvement because, in his opinion, a higher number of self-corrections in oral production may also indicate a more developed ability

---

that have been long identified as indicators (or events) that reflect fluent speakers not only in the L2 but also in the L1.

to monitor one's speech. This is somehow connected to proficiency level given that monitoring's one speech is a skill hard to develop by lower-level learners and is most frequent in the oral production of learners with higher levels of fluency, who probably also have a higher levels of proficiency.[2]

It is possible that due to the small number of participants included in his study, Lennon did not consider proficiency as an independent variable. Whether this is one of the reasons that could have possibly driven researchers interested in this field to stop considering level of proficiency as a factor influencing L2 fluency and not to consider possible correlations between these two factors of SLA is a mere presupposition. Unfortunately, the literature has avoided this issue and views proficiency and fluency as independent from each other (or at least, as one not being influential on the other). This dissertation is motivated in part by current lack of knowledge on the effect of proficiency on fluency; thus, the effect of proficiency will be examined for all fluency and psycholinguistic measures.

Occasionally, studies do not become influential for how they discuss a specific topic or what they support or demonstrate, but for what they suggest, advise or even for what they omit. At this early stage of the field, Lennon did not include in his study a cognitive analysis of second language fluency, but made a claim about the underlying cognitive mechanisms at work during L2 oral production by defining fluency as "an impression on the listener's part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently" (Lennon, 1990, p. 391). Ultimately, what Lennon investigated was the final product of these mechanisms (i.e., L2 speech delivered by his participants) before and after a stay abroad. Segalowitz (2010) also referred to this particular study and pointed out that Lennon defined

[2] In this line, Cucchiarini, Strik, and Boves (2002) examined the spontaneous speech of adult L2 learners of Dutch, reporting that *rate of speech* and *mean run length* between silent pauses explains the greatest amount of variance in fluency ratings of spontaneous speech made by expert raters.

fluency as "the listener's inference of the underlying 'cognitive fluency' that results from hearing more or less fluent speech" (p. 30).

Although Lennon (1990) is among the most cited studies from this decade, there were other studies 90's that accounted for fluency development, also focusing in the development occurring in SA contexts (Freed, 1995; Mehnert, 1998, Ortega, 1999; Riggenbach, 1991; Temple 1992; Towell, Hawking, and Bazergui, 1996). The growing interest in the field of L2 oral fluency research of the 1990's was expanded by far in the 2000's with the publication of studies that specifically addressed the relationship between second language fluency and its underlying cognitive abilities/mechanisms in different learning contexts (Link, Kroll, & Sunderman, 2009; O'Brien, Segalowitz, Freed, and Collentine, 2007; Segalowitz & Freed, 2004; Sunderman & Kroll, 2009).[3] These studies addressed the analysis of temporal variables and hesitation phenomena *a la* Lennon (1990) but have also added reaction time (i.e., RT) experiments and more complex statistical procedures to analyze their data. They also included more than one learning context (since traditionally, L2 fluency investigations such as Lennon (1990) and Towell et al. (1996) had been focused on the SA context only) such as the AH group and the less frequent *immersion* context (i.e., IM). In addition, the 2000's decade also saw the advancement of the use of technology to detect gaps of silence as well as syllables stress. For instance, Cucchiarini, Strik, and Boves (2000) developed automatic speech recognition software to account for L2 fluency. De Jong and Wempe (2009) developed a PRAAT script to detect syllable nuclei and measure speech rate directly from the sound file.[4] Although the use of this software is not very extended in the field, these are important contributions that will be enhanced in the

[3] A critical summary of these studies will be provided later in this chapter.

[4] Although this script saves time to the researcher by overcoming the need for transcriptions, it may also limit the type of fluency analysis, given that for any type of specific analysis of a second language learner's oral production, the researcher needs to direct access to a transcribed version of the text. This also helps to internalize the data and to track specific patterns in learners' productions.

future. Without doubt, there is a need for the inclusion of automaticity in the analysis process of oral fluency research to accelerate the process.

Two research groups that have stood out in this decade are what I call the Canadian School of L2 speech, lead by Tracy Derwing (University of Alberta) and Murray Munro (Simon Fraser University) on the one hand, and Norman Segalowitz (Concordia University) on the other. Derwing and Munro's research group has investigated the aspects that naïve listeners consider when they judge fluency (also known as *perceived fluency* (Skehan, 2003), such how self-corrections and filled pauses affect the perception of speech rate more than silent pauses (e.g., Derwing, Rossiter, Munro, & Thomson, 2004). Their findings also suggest a complex relationship between accent comprehensibility and speech rate and the way raters judge L2 speech (e.g., Derwing, Thomson, & Munro, 2006; Munro & Derwing, 2001). Segalowitz has published extensively over his career on L2 psycholinguistics (i.e., Favreau & Segalowitz, 1983; Segalowitz, O'Brien, & Poulsen, 1998; Segalowitz, Segalowitz & Wood, 1998; Segalowitz, 1997; Segalowitz & Gatbonton, 1995; Segalowitz, 2000), L2 fluency (i.e., Freed, Segalowitz, Dewey, 2004; Segalowitz and Freed, 2004), and the SLA of study abroad at large (i.e., Freed, Dewey, Segalowitz, & Halter, 2004; Segalowitz, Freed, Collentine, Lafford, Lazar, & Díaz-Campos, 2004). He also published in 2010 *Cognitive Bases of Second Language Fluency*, a book offering a state-of-the-art review of the research on second language fluency, particularly focusing on the development of cognitive mechanisms that underlie L2 oral production. This line of research motivates the current dissertation, as one of its primary objectives is to determine the effect of context of learning and proficiency on Spanish L2 fluency and cognitive abilities, with a special emphasis on reaction time measurements.

To predict future lines of research for the next decade of the 2010's is challenging, but

the goals will be to learn more about the specific aspects that promote fluency development in different learning contexts as well as to conceptualize fluency within a cognitive framework that accounts for fluency as a component of speech that is in constant flux and that is influenced by many linguistic and extralinguistic factors. In addition, more research will be necessary to understand fluency as a cognitive and oral skill that is deeply connected to speech perception, as well as to promote it as a core part of process of second language acquisition. This dissertation is motivated by the need to understand fluency from a cognitive perspective. It also has as one of its aims to advance the framework started by Lennon (1990) by investigating different measures of fluency (i.e., amounts of spoken time, syllables and filled pauses as well as rate of speech, rate of seconds per filled pause, and rate of syllables per filled pause) and also cognitive abilities (i.e., lexical access in Spanish and English, lexical retrieval in Spanish, and attention control in English) in the L1 and the L2. The next section summarizes work on one of the aspects that has encountered more discussion and disagreements in recent years: methodological approaches to fluency measures.

1.2.1.3. Methodological approaches on fluency

1.2.1.3.1 Operationalizing and measuring fluency

Amongst the most persistent drawbacks that second language fluency researchers face is the lack of agreement as to what components of speech should be measured to account for L2 fluency as well as what unit/s should be used from the large inventory of possibilities used in previous investigations. In essence, all these questions come down to the way each researcher operationalizes fluency and to the way each study is focused. There have been so many fluency measures implemented by researchers over the years that it is not difficult to find this single

aspect criticized in most studies on second language fluency that are currently being published.[5] Certainly, not all studies use the same measurements (or measure the same aspects involved in oral production), and for this reason it is hard to establish a specific number of measurements that would be sufficient to account for fluency. There is already agreement in the field that this is major methodological shortcoming (García-Amaya, 2008, 2009; Segalowitz, 2010), as this obscures comparisons between studies and provides more difficulties to the overall understanding of how second language fluency develops and which are the reasons behind the different outcomes that learners experience.

Over the years, researchers in second language fluency have agreed on the need to focus on *speed of speech delivery* and *hesitation phenomena* as two main tenets of L2 fluency because these aspects are among those that listeners use to judge someone's fluency. The most frequent measures used in second language fluency research are temporal variables of speech such as rate of speech and articulation rate, pausing phenomena including both filled pauses, and silent pauses, as well as lexicon related measures such as speech run between pauses, and amount of words spoken per turn (Griffiths, 1991), even though many other measurements have been suggested and some of them have more followers than others. Wood (2000) finds that speech run between pauses, which measures the number of words produced between pauses, is the most important discriminator for fluent vs. disfluent speech, because as he points out, highly fluent speakers tend to pause at sentence and clause junctures and produce longer runs, unlike less fluent speakers who produce shorter runs as they pause within clauses or sentences, thereby producing shorter runs. Recent studies seem to be in line with this particular assertion. For instance, García-Amaya (2010) showed that overall, L2 learners of Spanish produced less filled

[5] This is a fair critique rooted in the tradition of the scientific method but even more so in the tradition of the field of psychology, where the need to verify results through replication may be a more frequent practice than in the field of SLA.

pauses before and within different clause types after participating in 4-month SA and 7-week IM contexts.

In this dissertation, an important fluency measure is rate of speech. To date, fluency studies have calculated *rate of speech* in *words per minute* (Lennon, 1990; Freed, Segalowitz, & Dewey, 2004; Segalowitz & Freed, 2004), *words per second* (Möhle, 1984; Olynyk, D'Anglejan, & Sankoff, 1990; Raupach, 1980, 1984, 1987; Riggenbach, 1989; Sajavaara, 1987), *syllables per minute* (Towell, Hawkins, & Bazergui, 1996), and *syllables per second* (Temple, 1992, 2000, 2005; Binnenporte, Van Bael, den Os, & Boves, 2005; Czerwionka & García-Amaya, 2011; García-Amaya, 2008; 2009; García-Amaya & Filgueras-Gómez, 2010; Munro & Derwing 1995). All these measurements are meant to capture the same result, although different units are employed in the numerator and denominator. For instance, syllables per second is achieved by counting the number of syllables a learner produces in a selected fragment of speech (i.e., syllables is the numerator), and then dividing it by the total amount of seconds in which that particular fragment is produced (i.e., seconds is the denominator). The other rates included in this dissertation are second per filled pause and syllables per filled pause, which belong to hesitation phenomena.

With regard to words per minute, a number of authors have pointed out that this measurement should not be used in L2 research due the variable lengths of words, which could obscure developmental gains in L2 oral production (Freed, personal communication; Griffiths, 1991). Not all researchers, however, share this point of view. For example, Ejzenberg (1992) argues that words per minute is an accurate measure, as native speakers of Spanish and Portuguese add syllables to the original number of words in *Spain*, for example, which is normally produced as a disyllabic word (i.e., [es.pain] vs. [spein]). Although the example that

Ejzenberg (1992) includes is a well-known one, the number of syllables that Spanish-speaking learners of English could add to words starting with *s* in English is probably not as large as the number of syllables that go unaccounted for when words per minute is used to measure rate of speech. Taking into consideration the discrepancies in the literature, researchers have advocated for the use of syllables per second as the most accurate unit to account for the rate of speech of both NSs of Spanish and L2 learners of Spanish that are NSs of English (Binnenporte, Van Bael, den Os, & Boves, 2005; Czerwionka & García-Amaya, 2011; García-Amaya, 2008; 2009; García-Amaya & Filgueras-Gómez, 2010; Munro & Derwing 1995; Temple, 1992, 2000, 2005).

As was pointed out earlier, the methodological disagreements do not stop at the unit that should be used, but also at what to measure. For instance, some researchers include the exact time that a learner takes to produce a fragment, including everything from beginning to end (i.e., the time taken to produce words or syllables and the time to produce filled pauses such as *uh, um*), backchannels (e.g., *uh-huh, hmm*), and silent pauses. Other researchers measure *articulation rate* only, which accounts for the production of rate of speech excluding the time taken by pauses, or *pausing time*, which in many occasions, is considered another index of fluency. Studies including articulation rate include Chambers (1987) and Towell (1987), who justified the use of these measurements, by indicating that they were only concerned with speed of actual speech.

1.2.1.3.2. Lexical measurements

Studies addressing L2 fluency have also accounted for the amount of speech that learners are able to produce per turn. This measurement is normally calculated by summing the total number of words that a learner produces in a given fragment (e.g., Freed et al., 2004). The use of this measure is not as frequent as other more characteristic fluency measures such as *rate of*

speech or *articulation rate*. Other lexical measures include number of repetitions and number of repairs. Repetitions normally include words or fragments of words that are repeated by learners during oral production (e.g., *when, when I was, when I was doing my homework*). The term 'repair' is a more complex term that includes grammatical correction samples and other examples of online self-monitoring (e.g., *I didn't wanted to do, I didn't want to do this)*. Examples of repairs are not always clear-cut, and researchers in this field need to be detailed in their classification. Number of repetitions and repairs have been calculated by Lennon (1990) and Freed et al. (2004), and these measures have been used both to account for oral speech in qualitative and quantitative approaches to L2 fluency.

Other lexical items that have been used in the literature include lexical fillers (also called discourse markers in the discursive/pragmatic literature) such as *well, you know, actually, okay, well, actually,* etc. (Moder & Martinovic-Zic, 2004; Schiffrin, 1998) and quasi-lexical fillers (also called filled pauses, pause-fillers, and minimal vocalizations by Sacks, Schegloff, & Jefferson, 1974) such as *uh, uhm, hmm* (Griffiths, 1995; Temple, 1985). It is known that the use of these structures is not infrequent in NS speech. Although L1 speech certainly contains silent pauses, filled pauses, backchannels, repetitions, repairs, mispronunciations or even malapropisms (Clark & Fox-Tree, 2002), it has been discussed that in L2 speech, the overuse of these structures has been found to work as 'islands of safety' for language learners (García-Amaya, 2006). Indeed, lexical items such as lexical fillers (and quasi lexical fillers) are typically used by L2 learners when a word retrieval attempt fails or a planning issue arises during speech production. Certainly, these structures can have multiple roles and they should not be considered as mere 'storm shelters' for L2 learners. On the contrary, their role in L2 speech warrants further investigation. It is known that L2 learners can use these structures in order to hold the floor and

gain time while deciding how to continue, etc.

1.2.1.3.3. How much speech needs to be measured to account for fluency?

Researchers interested in L2 fluency do not always agree as to *which* measures should be used to account for fluency development, or as to *how* should one reach them. These discrepancies also extend to the minimum amount of speech/length of speech fragment necessary to account for an L2 learner's fluency level. Several authors have suggested that the most adequate procedure is to use a combination of comprehensive measures such as rate of speech and more specific measures such as articulation rate and pausing time (Segalowitz & Freed, 2004; Temple, 2005). This leads us to discussion of the details of the fluency analysis process.

Traditionally, L2 fluency research has required recording and transcription of speech from L2 learners. Although specialized software to measure rate of speech already exists (e.g., De Jong & Wempe, 2009), most researchers do this process manually, which depending on the extent of the sample and the number of participants can become a highly time-consuming task. Many researchers have encountered this setback with the immediate outcome of keeping the speech segments under analysis rather short. In fact, as García-Amaya (2008, 2009) points out the length of the fragments under study has been one of the main methodological disagreements of fluency studies in the existing literature. For instance, Freed (1995), Freed et al. (2004), Segalowitz and Freed (2004), and Temple (1992, 2005) calculated overall fluency measures by analyzing rather short fragments of speech. Freed (1995) studied fluency development in L2 learners of French by analyzing two 45-second speech samples, one recorded at the beginning of the semester and another one recorded at the end. Temple (1992, 2005) analyzed two samples of two minutes, each recorded at the beginning and end of the semester, in order to calculate an overall rate of speech for each of the 22 learners participating in her study. In the same fashion,

Freed et al. (2004) compared the production of 28 learners of French by analyzing two 1-minute samples taken from a pretest and posttest oral interview. Segalowitz and Freed (2004) analyzed two 2-minute segments from a pretest and two 2-minute segments in a posttest. Although all these studies were able to uncover differences in fluency measures by analyzing very short fragments of speech, they missed important aspects of interlanguage production, a point further pursued by García-Amaya (2008, 2009). In these studies García-Amaya critically analyzed previous second language research, by which measuring small samples of speech might have missed important aspects of the participants' speech production. In García-Amaya (2009), the elicitation task consisted of a controlled per topic questionnaire based on a Labovian sociolinguistic interview. For the analysis, the fifteen longest turns were analyzed for a variety of fluency measures including Spanish and English word production, rate of speech and rate of speech adjusted without reformulations and repetitions. The results indicated that there was an overall statistically significant difference among learner groups and that the ranges for each measurement indicated a significant amount of variation for different turns. This result led the later analysis of topic effect on rate of speech (García-Amaya, 2008), where a highly statistically significant effect was found for topic on rate speech, an aspect that had been overlooked by previous research.

As for elicitation tasks used in L2 fluency studies, they have mostly consisted of highly controlled tasks such as the Oral Proficiency Interview (OPI) (Freed et al., 2004; Segalowitz & Freed, 2004), spontaneous conversations (Temple, 1992, 2005), sociolinguistic interviews in the Labovian tradition (García-Amaya, 2008, 2009), picture descriptions tasks where participants asked to describe what is happening on a particular drawing (Ejzenberg, 1992; García-Amaya & Filgueras-Gómez, 2010) and video retells –a thorough explanation of this technique is included

in Chapter 2. The objective guiding these oral elicitation tasks is that participants talk as much as possible in order to collect a sample of their L2 oral production. Undoubtedly, each task renders different types of learner speech. For instance, the purpose of the OPI is that the participant performs a number of tasks to the best of his/her abilities. In this type of task the participant is aware that s/he is being scrutinized, and thus, depending on his/her performance a specific score is awarded, which in some cases may have professional and/or educational consequences.

The OPI differs from the sociolinguistic interview in that the latter was originally conceived by researchers interested in sociolinguistic variation to elicit different speech styles of a speaker's vernacular. In this type of interviews, the main purpose of the interviewer is to overcome the observer's paradox, or the task of eliciting a person's vernacular without necessarily belonging to the informant's speech community (Labov, 1966). Accordingly, the questions appearing in sociolinguistic interviews address past events, near-to-death experiences, future plans, and other common topics that a NS has no difficulty to talk about. Indeed, these types of elicitation tasks might have a number of drawbacks for fluency analysis if the topics covered are not appropriate to the specific learner or group of learners under investigation. As Lennon (2000) pointed out, certain aspects of the elicitation task can always be out of the control of the learner (e.g., the topic under discussion, the speech situation, the interlocutor, or the participant's mental state), and this can certainly influence oral production. Depending of the task, the minimal duration should be different. Thus, for a sociolinguistic interview, 15 to 25 minutes seems to be enough time for a learner to expand on ten to twenty questions. Other tasks such as picture or video descriptions may suffice with a few minutes depending on the complexity of the stimuli.

1.2.2 Contexts of learning

1.2.2.1 The origins of research on the SA experience

Carroll (1967) is the hallmark study on the effect of the SA learning context. This study analyzed, among other topics language proficiency for 2782 college seniors majoring in French, German, Italian, and Russian, reporting that experience abroad was a predictor of proficiency. In a summary of study abroad research from the 1960's, Freed (1995a) notes that back then only a number of studies denounced the lack of research focusing on the linguistic gains experienced by L2 learners participating in study abroad stays (Willis, Doble, Sakrayya, & Smithers, 1977). A response to this lack of interest in the linguistic gains by L2 learners in the SA context is found in three large-scale studies also reported by Freed (1995a). Dyson (1988) is a study on listening and speaking skills of 229 British students who studied in Spain, France and Germany for a year. The result was considerable improvement in both skills, particularly for the lower proficiency students (Freed 1995: 9). Meara (1994) includes self-reported data from 586 students who overall indicated that they felt that their oral skills did improve abroad, while their writing and reading skills did not. The third large-scale study is Coleman, Grotjahn, Klein-Braley, and Raatz (1994), which included 35,000 students from 100 institutions; the result was that after returning from an entire year abroad immersed in the target language, the latter slows down (Freed 1995, p. 110).

A number of smaller studies appeared in the 1980's that focused on fluency development by native speakers of German who were L2 French learners and native speaker of French who were L2 German learners (Moehle, 1984; Moehle & Raupach, 1983; Raupach, 1984, 1987). Interestingly, while the German learners of French did improve their rate of speech and length of time between utterances, their grammar accuracy and syntactic complexity did not improve. As

the interest in the SA context grew, a number of researchers decided to use the ACTFL Oral Proficiency Interview (OPI) to test oral proficiency development for students that had participated in a SA context (e.g., O'Connor, 1988 & Magnan, 1986 in French; Liskin-Gasparro, 1984; and Veguez, 1984 in Spanish; and Milleret, 1990 in Portuguese). This trend also continued in later studies (Freed, Segalowitz, & Dewey, 2004; Segalowitz & Freed, 2004). In general, these studies reported that learners that study abroad become more fluent and experience gains in oral proficiency. Very few studies however have investigated IM contexts, but all studies have concluded that learners do benefit and improve their speed of speech delivery as well as that they decrease their hesitation phenomena (D'Amico, 2010; García-Amaya, 2008, 2009, 2010; García-Amaya and Filgueras-Gómez, 2010).

1.2.2.2 Research on context of learning in Spanish

Previous studies on learning contexts in Spanish have focused either on the gains of SA contexts in isolation or on the differences between learners participating in SA contexts versus those that participate in the AH context. Researchers interested in L2 development in SA contexts have investigated college-sponsored programs in a foreign country where the target language is widely spoken; typically, the duration is either one or two academic semesters. One subtype of the SA setting is the overseas IM program that may have different durations varying from two weeks to two months. These programs are rather popular in Europe and the United States among high school and college students. Researchers have also investigated the AH context, which refers to classroom-based environments that are normally regarded as having a more traditional approach to language learning and teaching. The variety of teaching methodologies and styles found in the AH is highly varied and this might be the reason why the contexts of learning research up to date has focused on this context only as a source of

comparison for SA and IM programs. That is, it has not targeted specific groups of AH language classes such as honor and advanced classes where learners may be more driven and motivated than other L2 learners interested in completing a language requirement. This line of investigation may be of particular interest to those interested in individual differences and want to consider other aspects such as motivation, social factors, etc. Another AH context is the domestic IM setting (e.g., Freed, Segalowitz & Dewey, 2004 for L2 French). This type of program, like the overseas IM, has been characterized in the specialized literature as highly selective (i.e., prospective participants undergo rigorous screening process before they are accepted to participate) and intensive (i.e., learners will have several classes everyday). They are also costly, and learners normally have to pledge to different types of rules such as a language commitment.

As for research carried out on L2 Spanish, we could easily separate those investigations that have investigated one specific learning context from those that have compared different learning contexts such AH vs. SA and that have researched the following topics: grammatical abilities (e.g., Gunterman 1992a, 1993b; Isabelli 2001; Isabelli, 2004; Lafford & Ryan, 1995; López-Ortega, 2003), indicating that learners improve during their time abroad; cognitive styles (e.g., Hokanson, 2000), suggesting that introvert, extrovert, sensing and intuitive learners alike obtained similar oral and written gains; lexical acquisition (e.g., Ife, Vives Boix, & Meara, 2000), where it was pointed out that two semesters abroad provides more benefits than one; narrative abilities and fluency (e.g., Isabelli, 2001; Lord, 2009), which indicated that both SA learners considerably improve both their narrative abilities and oral production; pronunciation (e.g., Díaz-Campos, 2004, 2006; Simões, 1996), where it was proposed that learners improved their pronunciation abroad; and ethnographic research (e.g., Talburt & Stewart, 1999), where it

was found that some students suffer from racial and gender discrimination during their abroad stays.

As for research that compares SA to AH, the general consensus is that SA is more effective than AH in terms of oral production, cognitive development, grammatical abilities, and the lexicon. Díaz-Campos (2004), for instance, compared a group of L2 learners of Spanish in Alicante, Spain (i.e., SA group) with a group of L2 learners of Spanish in Colorado, US (i.e. AH group). Learners read a paragraph with 60 target words including different phonetic segments such as word-initial stops, intervocalic fricatives, word initial laterals, and palatal nasals. Although Díaz-Campos was unable to confirm a better pronunciation for either of the two groups of learners in this first study, the author, in a more recent investigation including a conversational task concluded better phonological abilities for the SA than for AH learners.[6] Stevens (2001) used both reading and storytelling tasks to test the development of Spanish pronunciation and he reported similar results. His findings indicated more gains for SA over AH learners for this particular linguistic skill.

Among the indisputable aspects that the specialized literature has reported as always benefiting from the SA abroad context, we find the topic of the present dissertation, that is, oral fluency. Although there are still many unanswered questions as to why and how this development actually occurs, and if actually it happens for all learners with comparable rates, the most current research concurs that fluency is probably one of skills from which learners benefit the most during the SA experience. As was previously mentioned, the term *fluency* stands for a large number of concepts that relate to the smoothness (or lack thereof) with which the listener perceives speech and with which the speaker engages in oral speech production. In this sense, the

[6] Lafford (2006) mentions that this result might be due to the nature of the reading task employed in the study, since it possibly triggered a more normative pronunciation and attention to form under conditions of controlled processing.

more ‘native-sounding’ attributes a learner has, the more fluent s/he will be considered (Möhle, 1984; Raupach 1984, 1987). Research in this area suggests that most students gain some type of global fluency by increasing the rate of speech and the length of time between pauses. They also abandon use of filled pauses and hesitation phenomena overall and acquire more native-sounding modifiers, formulae, and compensation strategies such as lengthening of sounds or discourse markers (García-Amaya, 2010).

Studies on fluency in L2 Spanish include DeKeyser (1986) and Segalowitz and Freed (2004), both of which indicate that SA groups surpass AM groups in fluency gains. More recently, García-Amaya (2009) interviewed intermediate learners of Spanish and reported that those learners that had participated in an SA program during the previous semester were able to speak with a faster rate of speech and with less filled pauses than the group that had remained at home. Moreover, he also compared the performance of these two groups with the performance of an overseas IM group of high school students that has spent 7 weeks in León, Spain and found that the IM learners, in a similar fashion than the L2 French learners in Freed et al. (2004), performed better than the SA group in all fluency measures.

DeKeyser (1986, 1991) and Lafford (2004) investigated the use of communication strategies, such as topic avoidance, message abandonment, and L1 literal translation. Both studies by DeKeyser compared SA and AH learner groups, and DeKeyser (1986) reported significant gains for the SA group on fluency but no significant differences between groups for grammar and communication strategies. The latter results were also replicated by DeKeyser (1991). Lafford (2004) found that the learners in the SA context consistently used fewer communication strategies than the AH group, a result that correlated negatively with the use of Spanish outside the classroom and with the host family.

In sum, the SA context has been shown to yield significant gains for L2 learners in terms of oral production, cognitive development, grammatical development, and the lexicon. Certainly, more research is necessary to strengthen this claim. Among the shortcomings of the studies mentioned above, it could be said that they have considered the performance of intermediate and advanced learners only. As was previously mentioned, researchers have proposed the existence of a *readiness threshold* that learners should attain before they embark on a SA program (Segalowitz & Freed 2004; Sunderman & Kroll, 2009). Thus, it may be that the current body of findings might be challenged if more learners with different levels of proficiency are investigated. In this regard, this dissertation goes beyond examination of which specific learning context is 'better'. Rather, it will provide insight on what learners (according to their level of proficiency) would benefit the most from which context and at what level. Additionally, this dissertation will go beyond L2 production, as it is typically formalized, and will investigate how underlying cognitive mechanisms interact with social factors (e.g., different learning contexts), and how individual L2 learners cope with it.

1.2.3 Background on cognitive abilities

1.2.3.1 L1 and L2 cognitive models of speech production

As Segalowitz (2010) points out, it is unfortunate that to date, no theoretical framework on second language fluency is yet to be developed (p. 7). For instance, if we compare current research on L2 fluency with theoretical approaches to spoken word recognition, we find that researchers in the latter field already count on specific models (i.e., TRACE, NAM, PARSYN) and can test these models, perfect them based on new findings, and most importantly, theorize how the processes of word recognition and speech perception occur in the human mind. A starting point for a theoretical approach to L2 fluency, however, lies in De Bot's (1992, 2003)

bilingual adaptation of Levelt's (1989, 1999) "blueprint" of what the act of speaking entails in the monolingual mind. Levelt's original model is a comprehensive model of L1 speech production, based on experimental research carried out during the previous 40 years focusing on psycholinguistic processes underlying both overt speech and error production. In this regard, Levelt's (1989) model of speech production, along with its updated version (Levelt, 1999), represent, still today, the most ample and comprehensive work to explain the underlying cognitive processes in human speech production.

De Bot's (1992, 2003) model of bilingual speech production is an attempt to adapt Levelt's (1989, 1999) model of monolingual speech production to L2 speech. Although, according to De Bot, Levelt's model does not require radical changes to be adapted to the L2, neither model addresses speech as a developmental and/or dynamic process. On the contrary, both models are steady and representative of 'snapshots' of the monolingual and bilingual production process in idealized speakers. Certainly, this structured view of the brain as a container of boxes, each one with one function, is not the most biological perspective that could be provided. Nonetheless, both models are important contributions, and although they are more successful in explaining how certain cognitive mechanisms underlie speech than in explaining in which way L1 and L2 speakers achieve this end, they represent a good start for the field to build a strong theoretical framework that considers the constant changes that L2 learners experience through the acquisition process.

Given this motivation, the next section is a brief summary of the main components shared by Levelt's (1989) and De Bot's (1992) models of L1 and L2 production and focus on the components/processes that were later reviewed in Levelt's (1999) and De Bot's (2003) models.

1.2.3.1.1 Levelt's monolingual speech production model (1989, 1999) & De Bot's bilingual speech production model (1992, 2003)

Levelt's (1989) model portrays the processes followed by monolingual speakers from the onset of an idea until its utterance. The model has five main modules/systems that account for the cognitive processes that underlie overt speech. The first module is the *conceptualizer*, in charge of converting a thought or idea generated by the speaker with the intention of being produced/uttered into a preverbal message. The preverbal message is a type of input that conveys the organization of a communicative intention but that does not include yet any words or specific propositional structure. The second system in the model is the *formulator*, in charge of coding the preverbal message with grammatical and phonological rules. The third system, the *articulator*, converts the speech plan into real speech, a process that is immediately followed by the processing and storage of the output of the formulator so that it can be fed back to the fourth system of the model, the *speech-comprehension system*. At this point, speech can be delivered and the oral production is itself fed to the auditory system, which will check for errors in overt speech. Finally, the fifth system involved in the model is the *monitoring system*, which supervises that the original communicative intention is actually expressed in speech.

In order to adapt the original L1 model of speech production to bilingual speech production, De Bot (1992) doubled the components of Levelt's model suggesting that bilinguals applied different principles to speak in the L1 and the L2 and that they have two conceptualizers, two formulators, two articulators, two speech comprehension systems and two auditory systems. However, in light of recent research suggesting that advanced bilinguals apply their L1-based principles of information when speaking their L2 (von Sutterheim, 2002), De Bot abandoned the 'doubling' of systems and in his more recent version of his bilingual model (De Bot, 2003), he rejected the existence of two conceptualizers. Other components that were originally thought to

exist in double form in the bilingual mind included the *formulator* and the *speech comprehension system*, even if the lexicons of each language were thought to be separate modules in the bilingual mind. Due to the inability of this first bilingual model (De Bot, 1992) to account for common language phenomena such as the activation of both languages simultaneously during code switching, or how this model could explain language encoding at two different levels of proficiency, De Bot (2003) abandoned the dual components structure arguing that their existence would impact both languages entirely. The only element in the system that remained unaltered in his updated version of the model is the *articulator*, a component whose hypothesis of existence is well-grounded by the fact that native-like prosody is hard to achieve by second language learners and that bilinguals normally have a 'foreign accent' in their second language because they possess only one articulator without a systematic division for the two languages and the L1 largely influences the L2. Finally, it is worth mentioning that one of the most innovative aspects of the revised model is the inclusion of a *language node* in charge of monitoring output and activating language-specific information on different levels of the process. Although De Bot has not developed this module in detail yet, its prospective role as part of the language learning system makes it very promising (Segalowitz, 2010).

As was previously indicated, the idea of updating the speech model was also applied by Levelt. In his revised version (i.e., Levelt, 1999), the main concepts were kept but a number of changes in the specifics of each component to the model were also added, partly to answer critics, and probably also to recognize ten years of specific advances in monolingual speech production and psycholinguistics at large. Instead of talking about specific components as he did in the 1989 model, in the new version Levelt details the functioning of each of the processes that are part of the speech production. An example of this is the coined term *macroplanning,* defined

as a process that involves the generation of ideas in the speaker's mind with a communicative intention. According to Levelt, it is at this first step in the process where a particular style (formal, informal) is coded.

In the second part of the communicative process, the speaker starts a new process to trim thoughts into concepts that can actually be expressed through words; Levelt named this process *microplanning*. Any thoughts that come to mind cannot be expressed with just one word. For instance, in order to express the thought of a man sitting in front of a canvas with a brush pointing at it, the speech production system would have to construct a proposition that involves the whole idea. The output of the *microplanning* process is the preverbal message (i.e., no words, just a train of ideas not yet developed), a component of the original speech production model that Levelt did not change in his updated version. The preverbal message then undergoes a process of grammatical encoding that requires lexical access to access the mental lexicon and more specifically, to the lemmas contained in it. A *lemma* is an abstract conceptual form of a word that represents the specific meaning of some entity but to which no specific sound has been assigned to yet.[7] These lemmas in the mental lexicon include groups of words, conventional expressions, and other lexicalized constructions. Once the preverbal message has been grammatically encoded the version that appears is the surface structure, where accessed lemmas surface to be assigned morphophonological codes by the phonological/phonetic system. The result is a phonological score, which is also the basis for the articulatory score, the latter contains information provided by the *syllabary* about gestural information, amplitude, pitch, and register, and that can start the motor activity necessary to produce real speech. Levelt's model is a very ambitious attempt to

[7] Some studies have shown that gender is retrieved from the word's phonological form (lexeme) and not from the lemma (Caramazza, 1997). Caramazza's Independent Network model does not assume a distinct level between the semantic and the phonological stages (thus, there is no lemma representation). In this model, syntactic information about the word is activated in the semantic or phonological level (so gender would be activated in the latter).

explain the very complex array of process involved in human speech.

Segalowitz (2010) develops a 'blueprint of the L2 learner' focused on second language fluency based on the models by Levelt (1989, 1999) and DeBot (1992). This is presented in the next section.

1.2.3.1.2 Segalowitz' 'blueprint' of the second language learner and its relationship to fluency

In an attempt to connect the development of L2 fluency for second language learners with the theoretical frameworks provided by the models by Levelt (1989, 1999) and DeBot (1992), Segalowitz (2010) has identified seven vulnerability points where L2 learners could encounter fluency issues. Starting at the top of the speech production process, there are different views on the specificity of the language selection process during macroplanning. According to Levelt at this point the speaker selects the register (i.e., formal/informal style) in which the thought or idea is going to be uttered, and DeBot (1992) along with Paradis (2004) believe that the L2 learner does not only select the register but also the language that is going to be used. Segalowitz disagrees, as he believes that these processes are non-language specific (p. 10) but does point out that certain topics or specific situations may require more microplanning than others, at which point, L2 learners may experience more disfluencies. This approach goes in line with the remarks of Lennon (1990, 2000) and García-Amaya (2008) who state that learners showed more disfluencies with topics that were less familiar to them (i.e., politics) than with the ones with which they had more familiarity (i.e., university life).

Second, microplanning may also present fluency difficulties for L2 learners who have trouble preparing their speech plans and its execution. This is another aspect where proficiency may play a role. Although in his book Segalowitz points out the importance of proficiency as a factor that needs to be considered in L2 fluency research, this was not included per se in the 2004

special volume on study abroad research (the authors did use OPI ratings but this test does not include proficiency). This dissertation has taken a different approach and has included a proficiency test to account for the effect of proficiency. Third, during the process of grammatical encoding, the process in which learners access their mental lexicon to access L1 and/or L2 lemmas in each language, learners may not be able to retrieve all the grammatical information requested by the surface structure; according to Segalowitz this may be another source for disfluencies. Certainly, there is a great deal of discussion as to how the process of lexical access and/or grammatical encoding takes place. Thus, it is still not clear whether someone that wants to say in L2 Spanish *table* thinks about the concept TABLE first and then retrieves *table* in English before accessing *mesa* in the L2, or if on the other hand, after thinking of the concept TABLE accesses and retrieves *mesa* automatically without interference from the L1. The experiments included in this dissertation show that large amounts of exposure to the L2 do have some effect on the L1 along with lexical access. These are limitations that Segalowitz points out for Levelt's (1989, 1999) and DeBot's (1992) models. Perhaps the biggest criticism of speech models is that they work like snapshots while L2 acquisition is dynamic and in constant change. It may be that a different approach such as the dynamics system approach (Larsen Freeman & Cameron, 2008) offers a more coherent answer. Lexical access is one of the processes within *grammatical encoding*, and this dissertation analyzes access to the mental lexicon through Spanish and English lexical access tasks, providing results collected *in situ* in an IM context. In addition, this dissertation analyzes L2 lexical retrieval (i.e., in Spanish), which also relates to the processes that take place between *grammatical encoding* and *overt speech*, that is, *morphophonological encoding*, *phonetic encoding* and *articulation*.

1.2.3.2 Cognitive abilities in different learning contexts

In recent years, the development of cognitive abilities and their relationship with L2 oral performance in different learning contexts has received a considerable amount of attention in the field of SLA. The studies that have investigated the interaction of cognitive abilities in different learning contexts have mainly investigated the following topics: cognitive abilities (e.g., lexical recognition, lexical access, attention control) and L2 fluency in an AH and SA context (Segalowitz & Freed, 2004); the role of phonological memory on fluency measures in an AH and SA context (O'Brien, Segalowitz, Freed & Collentine, 2006; 2007); the role of working memory resources in comprehension and production in an AH and SA context (Sunderman & Kroll, 2009); and the role of inhibitory control (Linck, Kroll and Sunderman, 2009).

1.2.3.2.1 Segalowitz & Freed (2004)

Segalowitz and Freed (2004) is one of the most influential studies in this particular subfield of SLA as it includes a number of never-before-tested measures of cognitive processing and analyzes possible correlations with oral production in two different learning contexts. The hypotheses brought to the field by Segalowitz and Freed (2004) have been highly influential for SA research. These authors go beyond explaining the development of specific L2 skill measures in a specific learning context and investigate the relationship between these skills and underlying psycholinguistic processes. Although their intention was to examine a possible connection between fluency and cognitive skills, the result opened a new field of inquiry about the relationship between fluency and cognitive abilities and their development in different learning contexts.

Segalowitz and Freed (2004) was based on the hypotheses of the skill-acquisition literature pertaining to the relationship of cognitive processing abilities and expert performance

(Ackerman, 1988); their proposal was that the degree of development of different cognitive variables might serve as *readiness* factors for oral gains. In order to exemplify this hypothesis, Segalowitz and Freed established a comparison between the need of the learner to process input appropriately in order to use it and what happens at the cognitive the learner when the learner is not able to produce output. In this second scenario, readiness factors, as they put it, have not reached the required cognitive threshold for the specific input to be processed; as a consequence, the learner might not experiment gains in oral production. According to Segalowitz and Freed, cognitive abilities might interact with learning experiences and affect L2 oral production. Thus, two cognitive abilities were included in their study, and each provided measures of speed (reaction time) and efficiency (accuracy) of *lexical access* and *attention control*, which had been previously discussed in contexts of L2 performance (DeKeyser, 2001; Segalowitz, O'Brien, & Poulsen, 1998; Segalowitz & Segalowitz, 1993), as well as in the context of L1 speaking and reading fluency (Levelt, 1989; Perfetti, 1985).

As for the experimental design, learners performed all cognitive tasks while they were in the United States, that is, before and after their experience abroad in Alicante, Spain. They performed a computerized semantic classification task, which required lexical access (i.e., word recognition) and in which participants had to make two-alternative forced-choice animacy judgments (e.g., *the boy* = living; *a boat* = nonliving). Attention control was also presented as a computerized test. Participants were asked to provide speeded responses to indicate which of three words appearing on the computer screen would match with another one (i.e., repeat condition) or would not match with the other two (i.e., shift condition). For example, participants saw a category such as *Where*?, and then matched that question with one of the following options: *above, inside, near, under.* Words included in the stimulus were quasi-randomized and

in no occasion did more than five similar judgment trials occur in sequence. The same tests were given at the beginning and end of the semester, and the experimental versions of the lexical access test yielded two measures in each language: speed of lexical access, as measured by reaction time; and automaticity of lexical access, measured by the coefficient of variation (CV).[8]

Results indicated that although learners in both contexts made gains over time in performance on the lexical access cognitive tests, there were no differences due to context. However, these results correspond to a comparison of what the participants experienced before and after they were immersed in their SA program and thus, do not truly represent how cognitive abilities had developed during the SA context exclusively. The authors did not point out this caveat and suggested that their results reveal that oral gains might depend on *cognitive readiness* to benefit from the learning opportunities available. For instance, their study was designed to compare oral gains while learning Spanish in different learning contexts, and it was established that more developed cognitive abilities would be related to the connection of words to meaning in an efficient way. This aspect would facilitate oral gains provided that better cognitive processing would foster communication. However, no direct evidence is presented for the existence of the *readiness* threshold.

The results of Segalowitz and Freed (2004) also showed statistically significant relationships between gains in oral performance and pretest levels of cognitive abilities, especially lexical access processing speed and automaticity (efficiency). In the pretests, both speed and efficiency of lexical access were positively related to oral fluency – that is, the degree to which the learner's speech was free of self-generated filled pauses. However, in the posttests,

[8] In psycholinguistic studies, the CV is used to calculate processing mechanisms and decide is they have reached a certain level of "automaticity" or "speedup". According to Segalowitz and Segalowitz (1997), "speedup" occurs when the mean reaction time and mean standard deviation of responses in an RT task decrease to the same degree, and "automatization" occurs when the mean standard deviation decreases more than mean RT. Thus, the CV decreases in the case of automatization while remaining unchanged in the case of speedup.

the efficiency of attention control had a negative relationship with speech rate. The authors suggests that learners more capable of shifting attention from one aspect of speaking to another employ part of their resources on that aspect of speech, thus slowing down their overall rate of speech. The data suggested that this latter effect may have been slightly more pronounced in the SA context, perhaps reflecting the greater demands and the greater need to self-monitor in a SA environment.

Given these motivations, the first goal of this dissertation it to investigate the linguistic outcomes offered by an *overseas* IM learning context in comparison to an AH program. Specifically, this dissertation investigates the amount of use of the target language for 27 L2 learners of Spanish participating in a six-week intensive overseas immersion program through the Indiana University Honors Program in Foreign Languages (i.e., IUHPFL) in León, Spain, as well as 29 L2 learners of Spanish enrolled in a fifth-semester Spanish class at Northern Illinois University in DeKalb, IL. An adapted version of the Language Contact Profile (Collentine & Freed, 2004) was used to understand differences in exposure and use of Spanish for participants in both learner groups. This analysis is very revealing of the realities of both contexts. Whereas the AH learners enrolled in one to three Spanish classes three days per week and spent the rest of their time exposed to their L1 (i.e., English), the IM learners attended five Spanish classes daily in addition to participating in a number of afternoon activities also conducted strictly in Spanish. Moreover, the learners in the IM program pledged to a no-English language commitment and were expected to use Spanish during the stay abroad (with very limited exceptions). In this particular overseas IM setting, the use of the L1 is sanctioned not only by the on-site instructors and host parents, but also by the students themselves. To date, researchers have yet to analyze

the effects of prolonged periods of L1 attenuation/inhibition on L2 learners' oral production in this context.

A second goal of this dissertation is to address the development of fluency (i.e., oral production performance) in different learning contexts taking into consideration the role of proficiency (defined as knowledge of Spanish grammar). Much has been said on the differences between the SA and AH learning contexts and the superiority of the former over the latter for fluency improvement (DeKeyser, 1986; Segalowitz & Freed, 2004), however, not many studies have focused on the distinction between shorter and more intensive overseas immersion programs (D'Amico, 2010; García-Amaya, 2008, 2009, 2010; García-Amaya & Filgueras-Gómez, 2010). For this dissertation fluency data were collected three times to investigate more closely specific developments of second language fluency in the IM context in comparison to the AH one. This will further allow to make claims about the optimal length of stay as a function of proficiency in IM contexts. Two overall fluency measures were calculated (i.e., speech delivery and hesitation phenomena), and these were based primarily on six production measurements.

Furthermore, recent studies on the overseas IM context in Spanish show that an 11-item proficiency test results in a ceiling effect for this particular group of learners (García-Amaya and Filgueras-Gómez, 2011; Geeslin et al., 2011, 2012). This dissertation has as one of its goals to account for differences in proficiency (i.e., low vs. high) with regard to fluency measures for the IM and AH groups by implementing a proficiency test with a larger number of test items.

A third goal is to investigate of the development of cognitive abilities for L2 learners of Spanish in the IM context. Given the effective foundation of the IUHPFL IM program, and the very specific characteristics that make it different from other contexts, particularly the intensive hours of formal instructions and the language pledge, it is important to understand what

cognitive development takes place not only in the L2 but also in the L1. For these reasons, this dissertation investigates the development of lexical access in Spanish and English words, lexical retrieval of Spanish words, and attention control with English stimuli. These cognitive abilities have been hypothesized to underlie oral production and as such, it was decided to provide a rigorous analysis of them. Indeed, the topics of cognitive abilities and L2 oral production have received a great deal of interest in recent years, particularly because the investigation of the correlation between these two aspects might shed light on when is the most ideal moment for an L2 learner to embark on an SA experience and take most advantage of the experience. In this regard, Segalowitz and Freed (2004) hypothesized the existence of a threshold of cognitive abilities whose attainment before the learner embarks on the abroad experience would result in higher gains in oral production skills. A more recent study by Sunderman and Kroll (2009) provides some evidence for a 'critical level of internal memory resources' that, according to these authors, is necessary to benefit fully from the SA experience. Unfortunately, the data provided in Sunderman and Kroll are not longitudinal, and therefore do not offer a complete picture to account for the existence of a threshold of cognitive abilities. Furthermore, in both studies data were collected when participants had returned to the United States and thus they may have missed what learners were truly experiencing at a cognitive level in both their L1 and L2 while they were surrounded by the L2. An important methodological goal of this dissertation was to collect data *in situ* in order to account for developmental changes while the IM learners were still fully engaged in their abroad program.

In sum, given the findings of previous research on different learning contexts, such as Freed, Segalowitz and Dewey (2004) and García-Amaya (2009), it is hypothesized that IM learners will become more fluent than the AH learners during the course of their respective

programs (i.e., 7 weeks abroad, a semester at large Midwestern institution). In other words, the IM learners will speak more (more syllables) and for more time (seconds), have a higher rate of speech, and produce less filled pauses. In addition, following the findings of Segalowitz and Freed (2004) for lexical access, it is also expected that the IM learners will become faster in their lexical access to Spanish and English words. Also, it is expected that as the IM learners develop their rate of speech, they will also improve their lexical retrieval abilities and not only access words in Spanish faster but also articulate them faster across time.

1.3 Outline of the dissertation

Given this background and motivation, this dissertation is outlined as follows. Chapter 2 provides information about the research methodology for the fluency and cognitive tasks. It also contains the demographic and background information for all participants under study (i.e., IM and AH learners), as obtained through the language contact profile. Briefly, the results show that there are statistically significant differences between the IM and AH learners on all four skills analyzed (i.e., speaking, listening, reading and writing) as well as a clear benefit for IM learners to interact in Spanish not only with native speakers but also with classmates in Spanish.

Chapter 3 contains the results and discussion of the pretest and posttest of the proficiency analysis, as well as the results and discussion of the fluency analysis of the video-retell task for the IM and AH learners. The most important findings of this chapter are that although both learner groups improve their level of proficiency from beginning to end of their respective learning experiences, IM learners overall improve more than AH learners. In addition, the IM group fluency showed a significant and steady rate of speech increase across time whereas the AH learners did not.

Chapter 4 includes the results of the cognitive tasks (i.e., lexical access, lexical retrieval and attention control) for the IM learners only. Generally, the findings were that while IM learners significantly increased their speed to access Spanish words (i.e., their L2) in their mental lexicon across time, they also significantly decreased their speed to access English words (i.e., their L1) by the end of the program. Faster RT times were also reported for the different conditions of the attention control task. However, no significant changes were seen for IM learners to retrieve Spanish words from their mental lexicon (i.e., to access and retrieve them) in a significantly faster fashion across time during the IM experience.

Finally, Chapter 5 offers a more in depth discussion of the results and a conclusion of this dissertation. It also discusses more general implications and outlines the needs for future research.

CHAPTER 2: BACKGROUND INFORMATION AND RESEARCH METHODS

2.1 Introduction

This chapter provides information on the research methodology that was devised for data collection and analysis for the current dissertation. Section 2.2 provides demographic and background information for the 27 immersion (IM) participants that performed the oral production and cognitive tasks, the 29 at-home (AH) participants that performed the oral production tasks only, and the language contact profile (LCP). Section 2.3 gives protocol information on the oral elicitation tasks and the cognitive tasks. Section 2.4 provides information on the IRB approval.

2.2 Participant information and Language Contact Profile

The design of the study included the participation of 56 native speakers of English who were L2 learners of Spanish: 27 IM and 29 AH. Learners in both groups shared an interest for Spanish language as shown by their average years of experience studying the language and by their career choices. All learners in the AH group were working towards a major or minor in Spanish at Northern Illinois University (NIU), and all IM learners were chosen among a larger pool of applicants who underwent a rigorous application process for admission into the Indiana University Honors Program in Foreign Languages (IUHPFL). Both learner groups were enrolled in a grammar course taught by the same instructor who used similar teaching materials and evaluation protocols. Thus, the only differences between both groups could be attributed as a product of the idiosyncrasies of each context (i.e., country/environment and the amount of classes per week, as well as the amount of self-generated and received Spanish native input, see below, sections 2.2.1 and 2.2.2). Table 2.1 provides demographic information obtained from the pre-test portion of the language contact profile (LCP) described in detail in the next section, as

well self-reported years of instruction in Spanish and English (also from the pre-test version of the LCP). Also, there is information on the amount of Spanish classes taken during the respective learning context.

Table 2.1 Demographic information and self-reported years of L1 (English) and L2 (Spanish) instruction

Group	Sex		Age	Years of Spanish instruction	Years of English instruction	Amount of Spanish classes
IM	21 F 7 M	Mean *SD*	17.04 0.45	4.41 2.04	11.48 0.78	5 0
AH	24 F 5 M	Mean *SD*	20.76 2.03	5.07 1.10	15.24 1.15	1.65 .61
t-tests and reports of significance			t(54)=9.325, p=.000	t(54)=1.448, p=.154	t(54)=14.603, p=.000	t(54)=28.294, p=.000

2.2.1 IM learners

The IM learners traveled to Spain on June 7, 2010 and resided in the city of León for a total of 7 weeks, until July 22, 2010. Classes met for 6 weeks, amounting to a total of 110 hours of formal instruction (in the seventh week no official classes were held). All learners enrolled in phonetics, grammar, conversation, literature, and culture classes. These learners also participated in extracurricular activities three days per week (choir or theater) and lived with host families. The data collection process for the IM group took place at three times during the experience abroad: during the first day of school upon arrival on June 9; at the end of the third week of classes, after 11 days of instruction on June 29; and at the end of the sixth week of classes, after another 11 days of instruction on July 19. A truly unique component of this immersion program was the 'language commitment' (or no-English rule), which stipulated that participants had to speak in the target language at all times during their stay in Spain. Program participants (and their parents) agreed to this language commitment as part of the application process, which

consisted of a Spanish proficiency test, reflective self-evaluations, a personal statement, letters of recommendation, and personal interviews.[9]

2.2.2 AH learners

The data for the AH learners were collected at NIU in Fall 2010 for two different sections of Advanced Spanish Grammar, a fifth-semester Spanish class which requires a total 37 hours classroom time throughout the academic semester. Out of a total of 29 AH learners, two were enrolled in grammar plus two other Spanish classes, 15 were enrolled in grammar and another Spanish class, and 12 were enrolled in the grammar class only. The first data collection date was August 31, 2010; the second was October 14 (12 class periods after the first data collection date), and the last on November 30 (also 12 class periods after the first data collection date). These participants performed the proficiency test, oral production task, cognitive tasks and the LCP in a computer lab available to them at the home institution.

2.2.3 Language contact profile

Traditionally, the use of a language contact profile enables the examination of any potential effects of traditionally relevant extralinguistic variables such as sex and socioeconomic class, as well as the characterization of the demographic makeup of learner groups. The language contact profile used in this dissertation is based in Freed, Dewey, Segalowitz and Halter (2004) and was chosen due to its completeness in questions about demographics, language-learning history, use of language in a given context, as well as contact with native speakers.

[9] As part of the self-evaluation, students were informed of the specific details of this program such as the language commitment and the requirement to attend intensive classes daily while abroad. This self-evaluation has proven to be beneficial to encourage motivated students to participate in the program. As part of this step students must also submit official school transcripts, a personal application with sections in Spanish, and letters of recommendations from teachers who are in a position to assess the strengths and weaknesses of each candidate. Specific information on the application process is summarized Appendix B.

The language contact profile is divided into a pretest and a posttest. The pretest version has two parts. Part 1 contains questions about gender, age, and experience with languages in formal educational settings, etc. Part 2 contains questions about prior use of and exposure to Spanish. As for the posttest version of the language contact profile, the original version by Freed et al. (2004) contained nine main questions with four sub-questions for a total of 36. These questions pertained to every aspect of the abroad experience relevant for the researcher from the living arrangements while abroad to the amount of Spanish used and classroom contact time. As recommended by Freed et al. (2004), I gave both groups the pretest version as the last step of the data collection process at Time 1, and the posttest version as the last step of the data collection process at Time 3. Since Freed et al.'s (2004) version of the language contact profile was originally intended for college students studying Spanish in an abroad setting, the version given to the AH and IM groups had to be modified based on the specifics of each group. Thus, all information pertaining to university or college life was modified. The pretest and posttest of the language contact profile can be found in Appendix A.

2.2.4 Comparison of self-reported data in the posttest version of the LCP

This section presents the results of self-reported data from the posttest version of the LCP administered to the IM and AH leaners. As was previously mentioned, the questions appearing in the posttest version of the LCP pertained to the particular aspects of the IM and AH experience including information about housing as well as detailed questions about use of the target language. The posttest version of the LCP was considerably longer than the first part and was administered to all participants at the end of their learning experiences.

Figure 2.1 shows the overall difference between self-reported time speaking Spanish for the AH and IM learners during their respective learning contexts. The most striking difference

between the two groups concerns the responses on self-reported time speaking Spanish, where overall, the IM group's responses indicated large differences in comparison to the AH group in all categories. Although the statistical analysis indicated that there was a highly significant difference between groups on the total spoken time, $t(54) = 16.663$, $p=.000$, there were also significant differences between the IM and AH group for speaking Spanish: to classmates, $t(54) = 18.345$, $p=.000$; with non-native speakers of Spanish, $t(54) = 13.143$, $p=.000$; with native speakers of Spanish, $t(54) = 12.320$, $p=.000$; with roommate, host family, or in the dorm, $t(54) = 10.879$, $p=.000$; outside class for extended conversations with host family and roommates, $t(54) = 9.330$, $p=.000$; outside class for superficial or brief exchanges, $t(54) = 7.230$, $p=.000$; while using things taught in the classroom with native or fluent speakers outside class, $t(54) = 6.246$, $p=.000$; with instructor $t(54) = 5.143$, $p=.000$; with friends, $t(54) = 2.857$, $p=.006$; outside class to clarify classroom-related work, $t(54) = 3.613$, $p=.001$; outside class to obtain directions or info, $t(54) = 4.638$, $p=.000$; when taking things learned outside class back to the classroom for question or discussion, $t(54) = 4.358$, $p=.000$; with strangers, $t(54) = 3.302$, $p=.002$; and with service personnel, $t(54) = 5.002$, $p=.000$. The only category of self-reported time speaking Spanish where there was not a significant difference between the IM and AH groups included *other*, $t(54) = 1.248$, $p=.217$. Also, there were no statistical significant differences between AH and IM learners for time spent speaking a language other than English or Spanish, $t(54) = .904$, $p=.370$.

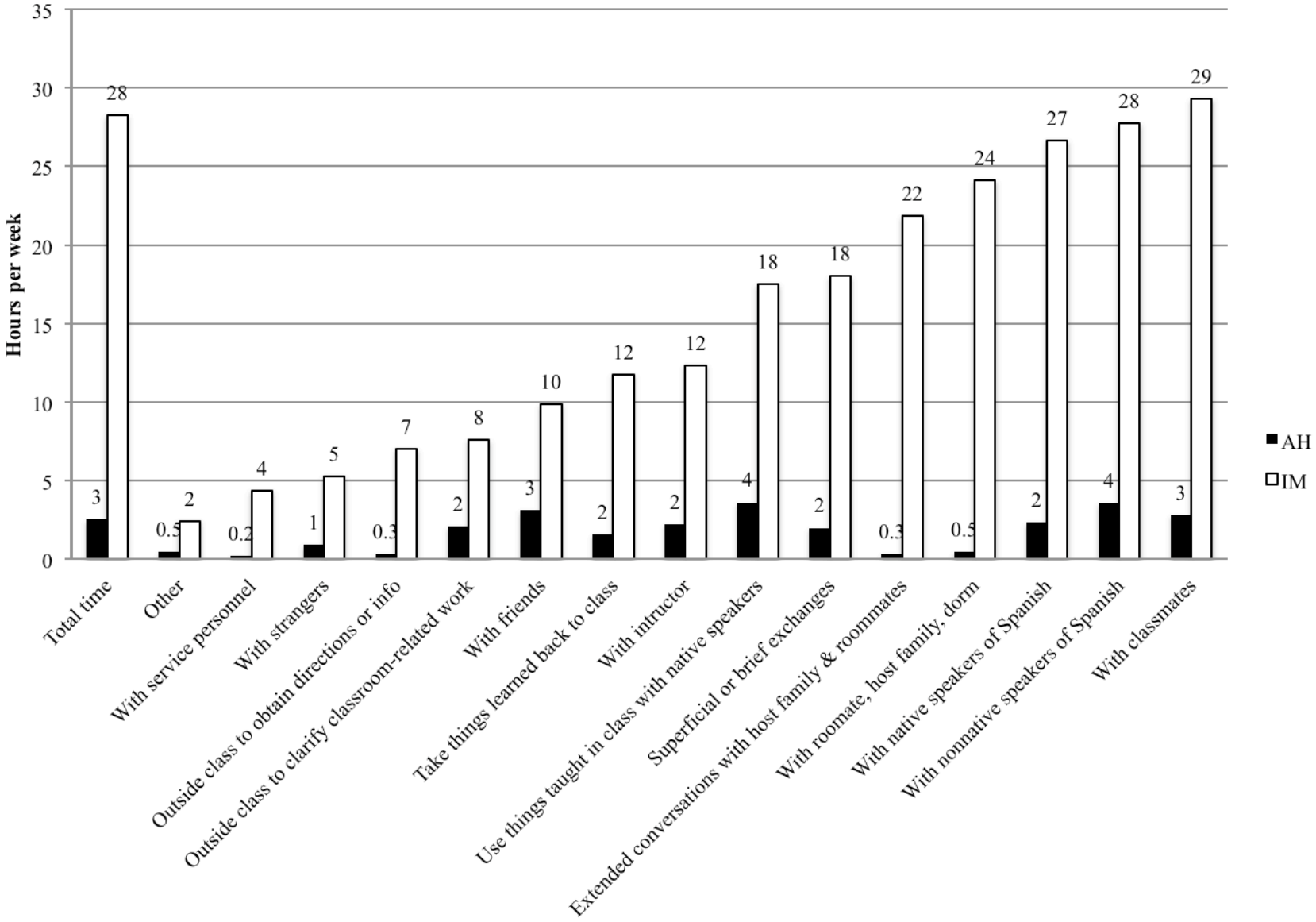

Figure 2.1. IM and AH self-reported time spent speaking Spanish in different situations

Figure 2.2 shows the average self-reported time spent writing in Spanish outside of class, based on the posttest version of the LCP. IM learners spent overall more time writing in Spanish than AH learners, and the analysis revealed a significant difference between both groups, $t(54) = 4.475$, $p=.000$. The same applied for writing homework assignments, $t(54) = 4.228$, $p=.000$, for writing emails in Spanish outside class, $t(54) = 2.878$, $p=.006$, for writing personal notes or letters in Spanish outside of class, $t(54) = 3.180$, $p=.002$, and also for filling out forms or questionnaires in Spanish outside of class, $t(54) = 3.431$, $p=.001$.

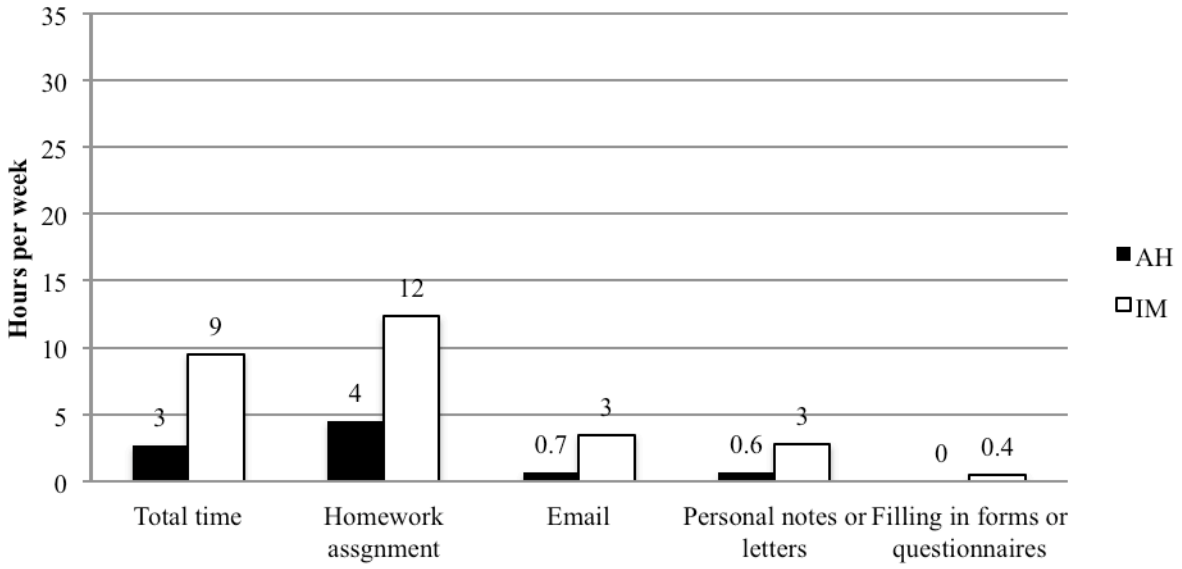

Figure 2.2. AH and IM self-reported time spent writing in Spanish outside of class

With regard to self-reported time spent reading in Spanish outside the classroom, as figure 2.3 shows, IM learners spent more time reading in the L2 than AH learners during their respective language learning environments, $t(54) = 2.612$, $p=.012$. The statistical analysis also indicated that there were a number of significant differences between groups for time spent reading schedules, $t(54) = 4.991$, $p=.000$, reading email or Internet websites in Spanish, $t(54) = 2.760$, $p=.008$, reading novels, $t(54) = 2.336$, $p=.023$, and reading newspapers, $t(54) = 2.608$, $p=.012$. No significant differences were found for reading magazines, $t(54) = 1.452$, $p=.152$. One finding worth clarification is that IM learners self-reported spending 8.18 hours per week reading

schedules, announcements and menus outside of class, versus only 1 hour reported by the AH learners. Although all other average numeric differences between the two groups in Figure 2.3 seem reasonable, it may be that for this particular category IM learners added extra time due to arrangements in their classroom facilities. Specifically, IM learners ate lunch during class days at a local restaurant that changed its lunch menu everyday and required that the students choose their meal preference before their arrival at the restaurant. Typically their instructors presented them with the menu options as well with the ingredients for each dish. This procedure lasted between 5 and 20 minutes everyday.

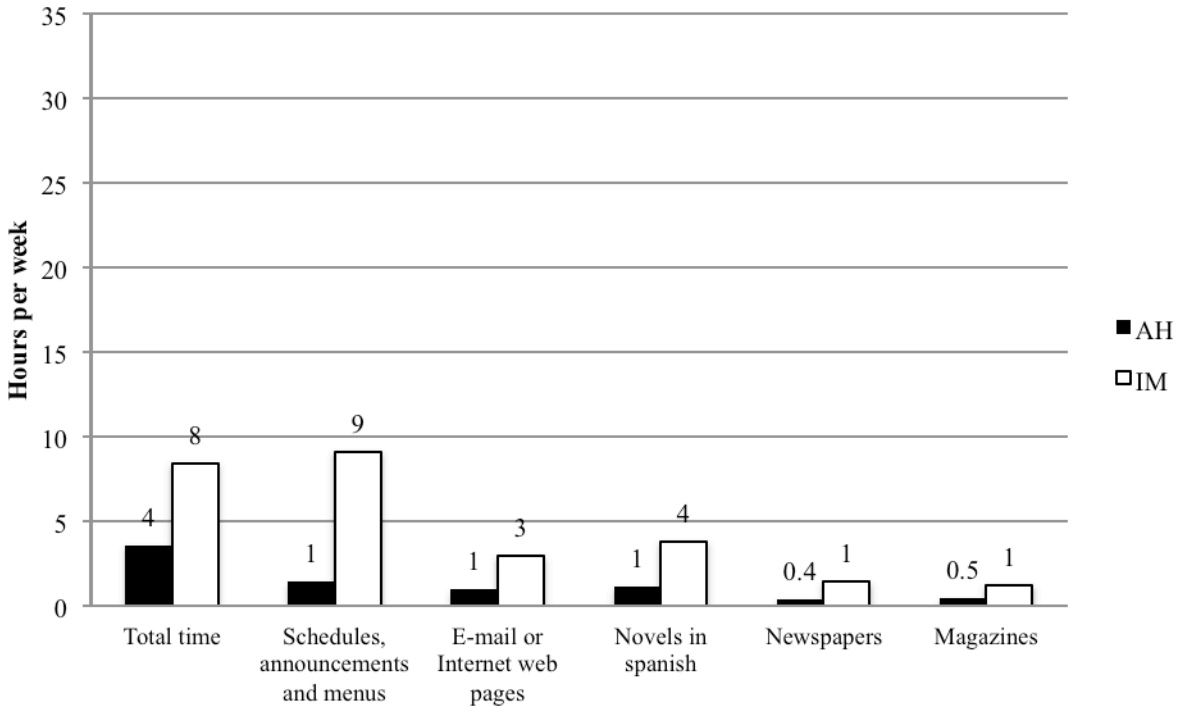

Figure 2.3. AH and IM self-reported time spent reading in Spanish outside of class

Listening was another skill included in the questionnaire. Figure 2.4 shows that IM learners self-reported more time spent listening to Spanish outside of class than the AH group, $t(54) = 11.181$, $p=.000$, listening to television and radio, $t(54) = 6.435$, $p=.000$, listening to Spanish songs, $t(54) = 6.128$, $p=.000$, trying to catch other people's conversations, $t(54) = 3.116$, $p=.003$, and also listening to Spanish movies or videos, $t(54) = 4.656$,

p=.000.

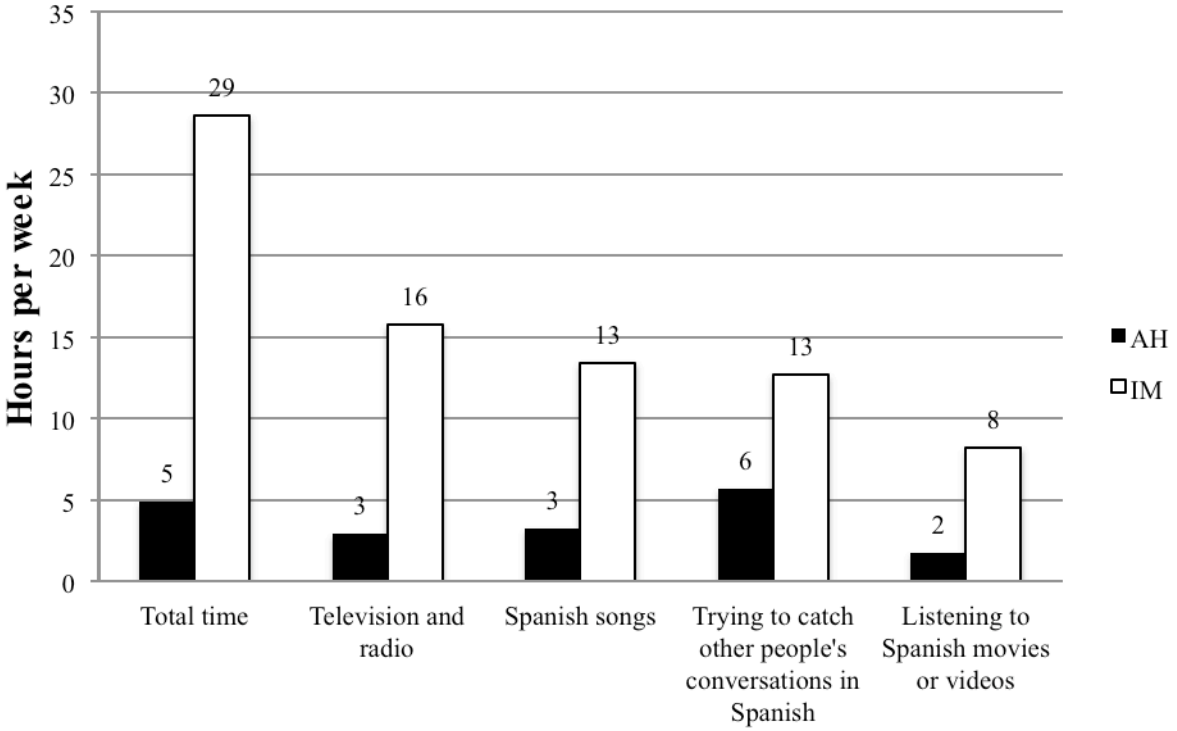

Figure 2.4. AH and IM self-reported time spent listening in Spanish outside of class

Figure 2.5 shows that the amount of self-reported *English* spoken by the IM learners was remarkably minimal (i.e., 20 minutes per week outside class), which we can attribute to the learning context itself as well as to the language commitment. The amount of English spoken by the AH learners reflects an expected amount of time that American college students would spend speaking their native language outside of class (32 hours per week), and the statistical analysis indicated that there was a highly significant difference between the IM and AH groups, $t(54) = 18.411$, $p=.000$. There were also significant differences for time spent speaking English to nonnative speakers of Spanish, $t(54) = 4.256$, $p=.000$, and for time spent speaking English to native or fluent speakers of Spanish, $t(54) = 2.452$, $p=.017$. Clearly, the amount of L1 used in the two learning contexts was an important differentiating variable in the analysis: IM learners

almost never used English outside of the classroom, whereas AH learners almost always spoke English outside of the classroom.

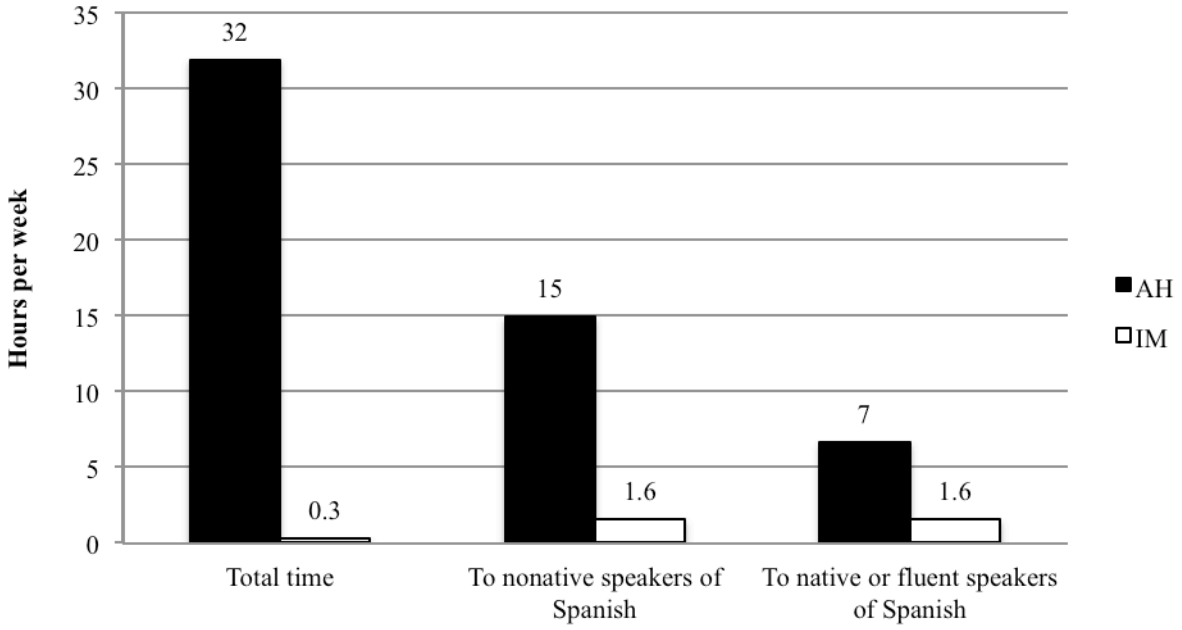

Figure 2.5. AH and IM self-reported time spent speaking in English outside of class

When asked about their use of English during the program, all IM learners indicated that they wrote in English more than they spoke it. The self-reported data for time spent writing in English for IM and AH learners are represented in Figure 2.6. Overall, IM learners spent 2 hours per week writing English, and AH learners spent 14 hours per week writing English. It is worth mentioning that the IUHPFL allows student participants a maximum one hour per week to correspond with family and friends online. The statistical analysis indicated that the difference between both groups was highly significant, $t(54) = 4.208$, $p=.000$.

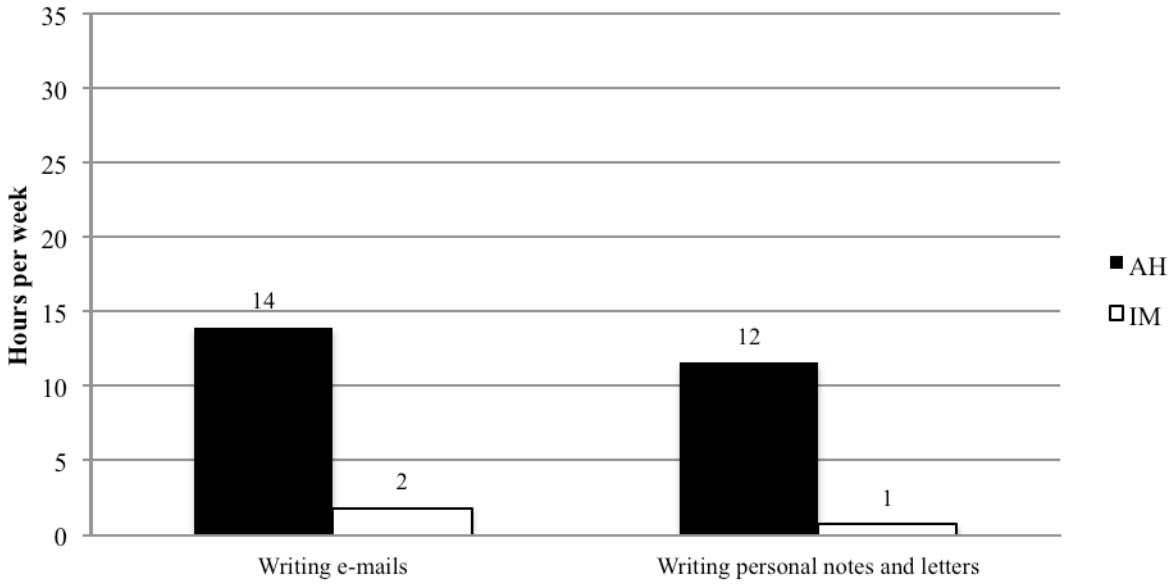

Figure 2.6. AH and IM self-reported time writing in English

In addition, the IM learners also reported minimal amounts of time spent reading, listening or watching movies in English. These averages were considerably higher for all AH learners, as shown in Figure 2.7, and again, the statistical analysis indicated a highly significant difference between the two groups, $t(54) = 4.015$, $p=.000$.

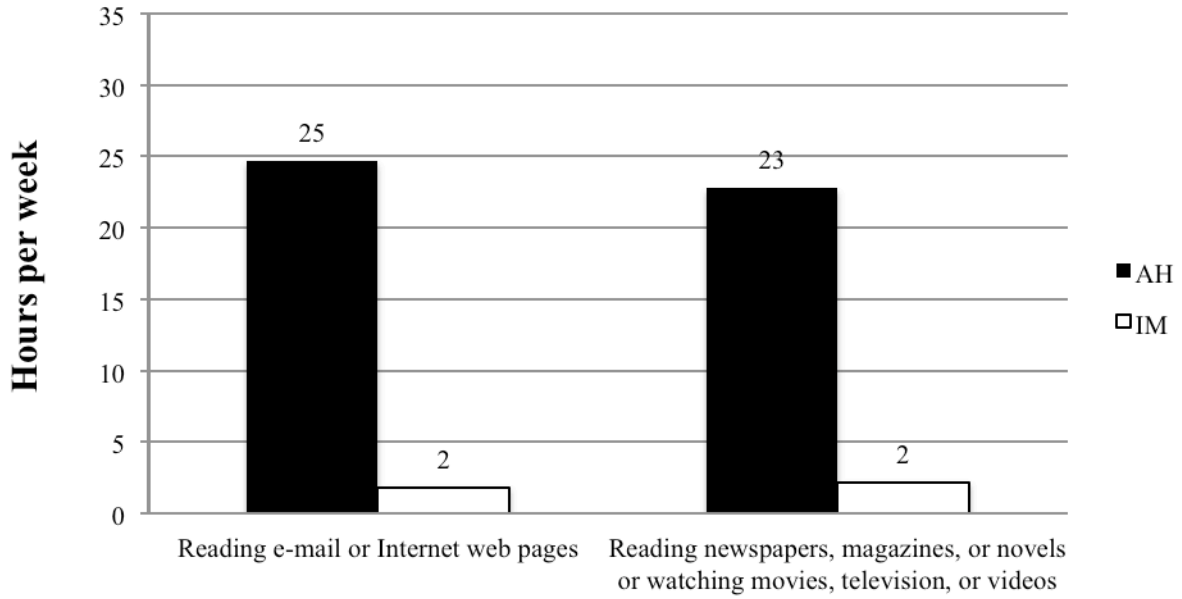

Figure 2.7. AH and IM self-reported time spent reading and listening in English

In sum, these self-reported data provide a better insight into L1 and L2 language use for each learner group. More specifically, this section on self-reported time spent speaking, writing, reading and listening Spanish and English by the AH and IM learners has provided quantitative evidence, that overall, IM learners used Spanish almost exclusively and resorted to English on limited occasions during their immersion experience. On the contrary, AH learners used much less Spanish outside the language classroom. They reported an overall 3 to 5 hours per week for speaking, reading, listening and writing in the target language.

2.3 Data elicitation tasks

All participants included in the current dissertation performed seven data collection tasks: a grammar test; a language contact profile (LCP); a video retell; a Spanish semantic classification task (Spanish SCT); an English semantic classification task (English SCT); an English attention control task (ACT), and a Spanish picture-naming task (PNT). For each group there were three data collection time points, and Table 2.2 indicates the timeline that each group followed along with the specific dates on which data were collected for each group. Participants

completed three versions of the video retell (i.e., Time 1, Time 2 and Time 3); three versions of the Spanish semantic classification task (i.e., Time 1, Time 2 and Time 3); two versions of the English semantic classification task (i.e., Time 1 and Time 3); two versions of the English attention control task (i.e., Time 1 and Time 3); and three versions of the Spanish picture-naming task (i.e., Time 1, Time 2 and Time 3). All method, stimuli and procedures are described in sections 2.3.1, 2.3.2, and 2.3.3.

Table 2.2. Data collection dates for both IM and AH group for both oral and cognitive data.

	Time 1	**Time 2**	**Time 3**
IM group	June 9	June 29	July 19
AH group	August 31	October 14	November 30
Tasks	Proficiency test Video retell English ACT Spanish SCT English SCT Spanish PNT LCP	Video retell Spanish SCT Spanish PNT	Proficiency test Video retell English ACT Spanish SCT English SCT Spanish PNT LCP

2.3.1 Proficiency test

All participants completed a 45-item proficiency test. The proficiency test was administered to the L2 learners as a pre-test immediately following their arrival to Spain and again as a post-test following completion of the final set of data elicitation instruments at the beginning of the last week in Spain. A similar 11-item multiple-choice version of this proficiency test was previously found to be a statistically reliable indicator of proficiency (Woolsey, 2006). This proficiency test covered a range of grammatical structures generally included in formal instruction of Spanish. However, three recent studies that used this proficiency test with IM learners similar to the ones included in the current dissertation (i.e., García-Amaya & Filgueras-Gómez, 2010; Geeslin, García-Amaya, Barker, Henriksen, & Killam,

2010, 2012) uncovered a ceiling effect for IM learners. In order to avoid this limitation, 34 more grammar items were added. The proficiency test can be found in Appendix C.

2.3.2 Oral production task

2.3.2.1 Video retell

Video retelling is a naturalistic data collection procedure in which stories can be presented in written, picture or video mode. This elicitation technique, in which learners are shown a story that needs to be retold later, has been a widely used elicitation technique in SLA research to investigate the foreground-background relationships in ESL (e.g., Tomlin, 1984) as well as to elicit L2 tense and aspect features. Bardovi-Harlig (2000) points out a number of benefits of 'story retellings'. First, in retellings the researcher knows the sequence of events. Second, story-retells can be compared across learners. Third, retell tasks may encourage some learners to produce longer samples than they would otherwise. Fourth, the content of stories may be manipulated by the researcher. In addition, story retellings, particularly in picture and video mode, also place higher demands on memory than other types of L2 elicitation techniques such as oral interviews. This data elicitation methodology has been implemented in SLA research in many learning contexts and also for many proficiency levels (e.g., Godfrey, 1980; Gass, Mackey, Álvarez-Torres, and Fernández-García; 1999; Hyltenstan, 1988, 1992; Larsen-Freeman, 1975; Salaberry, 1998; Sanz, 1997; VanPatten & Sanz, 1995).

In this dissertation, IM and AH learners retold videos of Simon Tofield's 'Simon's Cat'. There were two different videos at each data collection time for a total six videos over the three data collection times. All Simon's Cat videos share a similar theme about a mischievous cat. As Table 2.3 shows, each video has a variable duration between 37 and 149 seconds. It was decided that all participants watch two different clips at each data collection time, as researchers have

pointed out the effects of task repetition on measures of oral production (e.g., Bygate, 1996; Robinson, 2001).[10] The rationale for using different videos at each recording time was to avoid an improvement in fluency measures due to task repetition effects over the course of the language-learning context.

Table 2.3. Simon's Cat videos duration at each time point

	Time 1		Time 2		Time 3	
Videos	Video 1	Video 2	Video 1	Video 2	Video 1	Video 2
Video titles	"Fly Guy"	"Hot Spot"	"Let Me In"	"Snow Business"	"Cat Man Do"	"TV Dinner"
Individual video timing	0:06-2:23	0:06-0:43	0:06-1:52	0:06-2:18	0:06-1:38	0:06-2:35
Individual video duration	137 sec	37 sec	106 sec	132 sec	92 sec	149 sec
Total amount of time for videos 1 and 2	174 sec		238 sec		241 sec	

All six Simon's Cat's videos were downloaded from the Simon's Cat website and embedded into three separate Microsoft PowerPoint presentations: two videos for Time 1; two videos for Time 2; and two videos for Time 3.[11] The PowerPoint instructions indicated that participants retell what they saw immediately after watching each video. Participants were instructed to wait until each video ended to begin their retells. The digital recorders were never stopped during the retells, and in the PowerPoint presentation there was information directing learners to call an assistant or the researcher as soon as all oral data had been recorded. Participants were allowed to spend as much time as they wanted to complete their oral retellings.

IM learners viewed the PowerPoint presentations in laptop computers in León, Spain, and AH learners viewed the presentations on desktop computers in a language lab at their home

[10] Bygate (1996) found increases in accuracy, complexity and fluency as a result of practice when having participants repeat the same task. Robinson (2001) used repeated versions of a task that differed in complexity and argued that complexity differentials would be a more powerful influence on production than repetition of task versions.

[11] Samples of Simon's Cat videos can be found online at http://www.youtube.com/watch?v=Tuf61OjvoPQ&feature=fvhl and http://www.youtube.com/watch?v=w0ffwDYo00Q&NR=1.

institution. For the IM group, data were recorded using a PMD620 Marantz or DR-07 Tascam digital audio-recorder and a Shure head-mounted microphone. For the AH group Tandberg Educational MB Quart OEM's microphones were used with the Sanako recording system.

2.3.2.2 Analysis protocol of the oral production task

As was explained in Chapter 1, although there are a number of methodological issues that still need further consideration, researchers have gone to great lengths to analyze the subtleties of temporal variables in fluency measures and have included repetitions, reformulations, articulation rate, speech rate, filled pauses per second, etc. both in research on L2 Spanish fluency (Freed, 1995; Bartegui 1997; Segalowitz & Freed, 2004 among others) as well as in other second languages such as French (Freed, Segalowitz, & Dewey, 2004) and Portuguese (Eizenberg, 1992). These oral data were later analyzed for a series of fluency measurements under two major *overall* measures to capture speed of speech delivery (measures 1, 2, and 3) and hesitation phenomena (measures 4, 5 and 6) because these two *overall* measurements have been shown as a reliable index of fluency development. Speed of speech delivery measures include: 1) total amount of speech spoken (measured in syllables), 2) total amount of spoken time (measured in seconds), and 3) rate of speech in syllables per second. The hesitation phenomena include 4) amount of filled pauses produced, 5) seconds per filled pause, and 6) syllables per filled pause. . Other individual measurements that were also taken into account include: amount of speech spoken in the L2 (measured in syllables); amount of time spoken in the L2 (measured in seconds); production of filled pauses; rate of filled pauses per second; and rate of syllables per filled pause. Specific information of each measurement follows in the rest of this section.

First, the *amount of speech spoken in the L2* was measured in syllables, and its purpose is to account for the total number of syllables that learners produce in the L2. It is measured by

counting the number of syllables produced by the learner in the target language. This measurement has been used by García-Amaya (2009) for L2 Spanish (L1 English) and Freed, Segalowitz, and Dewey (2004) for L2 French (L1 English).

Second, the *amount of time spoken in the L2* was measured in seconds. This measurement accounts for the total amount of time taken by the participant to retell the videos. It has been used previously by García-Amaya (2009) for L2 Spanish (L1 English) and by Freed, Segalowitz and Dewey (2004) for L2 French (L1 English), among others.

Third, *rate of speech* was measured in syllables per second and accounts for speed of production in the target language. In this comprehensive fluency measure the numerator is all syllables from L2 words, and the denominator is the entire time it takes the learner to produce the turn under analysis, in seconds. This measurement has been used in: García-Amaya (2008, 2009) for L2 Spanish (L1 English); Temple (1992) for L2 French (L1 English); Eizenberg (1992) for L2 Portuguese (L1 English); Freed (1995) for L2 French (L1 English); and Lennon (1990) for L2 English (L1 German).

Fourth, *filled pauses* were counted to account for vocalizations such as *uhs* and *uhms* in the oral production of L2 learners. This measurement has been used extensively in L2 studies of NSs of English learning either French or Spanish as a second language: Freed (1995) for L2 French, and García-Amaya (2009, 2010), and García-Amaya and Filgueras-Gómez (2010) for L2 Spanish.

Fifth, *rate of seconds per filled pause* indicates the number of seconds that pass until a learner produces a filled pause. This is measured by dividing the number of seconds in a turn by the total number of filled pauses that appear in it. It was noted that *rate of filled pauses per second*, a measure previously used in García-Amaya (2009) for L2 Spanish (L1 English),

provided a number smaller than the unit and that it was harder to compare this rate among learners.

Sixth, *rate of syllables per filled pause* shows the number of syllables between filled pauses. This is measured by dividing the number of syllables in a turn by the total number of filled pauses that appear in it. This measure was used in García-Amaya (2010) for L2 Spanish (L1 English) and provided another frame of comparison between different learner groups. In sum, three measures contributed to quantification of Rate of speech, and three measures contributed to quantification of Filled pauses.

Table 2.4 summarizes these fluency measures and provides relevant background references.

Table 2.4. Fluency measures used in this dissertation (also used in previous studies)

	Temporal variables	Unit	Previous studies using this or a similar measurement
Rate of speech delivery	Rate of speech	Syllables / second	García-Amaya (2008, 2009) for L2 Spanish (L1 English); Temple (1992) for L2 French (L1 English); Ejzenberg (1992) for L2 Portuguese (L1 English); Freed (1995) for L2 French (L1 English); Lennon (1990) for L2 English (L1 German);
	Amount of speech spoken in the L1	Number of syllables or words produced	García-Amaya (2009) for L2 Spanish (L1 English); Freed, Segalowitz and Dewey (2004) for L2 French (L2 English)
	Amount of speech spoken in the L2	Number of syllables or words produced	García-Amaya (2009) measures syllables and words for L2 Spanish (L1 English); Freed, Segalowitz and Dewey (2004) measured words for L2 French (L1 English)
Hesitation phenomena	Filled pauses	Number of filled pauses	Freed (1995) for L2 French (L1 English); García-Amaya (2010), García-Amaya and Filgueras-Gómez (2010) for L2 Spanish (L1 English)
	Filled pauses seconds / filled pause	Number of seconds / filled pause	García-Amaya (2009) for L2 Spanish (L1 English)
	Rate syllables / filled pause	Number of syllables / filled pause	Garcia-Amaya (2010) for L2 Spanish (L1 English)

2.3.3 Cognitive tasks

The four remaining psycholinguistic tasks were designed to examine different cognitive abilities that previous research had suggested underlie oral production. The cognitive tasks were: a Spanish semantic classification task (which measures L2 lexical access); an English semantic classification task (which measures L1 lexical access); an attention control; and a picture-naming task (which measures working memory resources for lexical access and lexical production). Each of these tasks measures accuracy (% or correct responses) and reaction time (i.e., RT) in milliseconds. RT is a measurement that indicates the elapsed time between the presentation of the chosen stimulus and the response given by the participant. By indicating how fast a participant mentally performs a given task, RT data provide an index of processing speed.

Only the IM learners completed the cognitive tasks. The four cognitive tasks that were included in this dissertation (i.e., semantic classification task in Spanish, semantic classification task in English, picture-naming task in Spanish, attention control task in English) were programmed with the software DMDX (Forster and Forster, 2003). This specialized software required the preparation of a script (i.e., .rtf file), which contained the stimuli that were presented to the participants. Each file contained detailed specifications about the presentation of each task, fonts, colors, instructions for the participants, amount of time that each slide and stimuli were presented, as well information for correct and incorrect responses. Once participants performed each task, their responses (i.e., RT measured in milliseconds, and whether the response was correct or incorrect) were registered in an .azk file and then recoded for input into SPSS. In the following subsections, each cognitive task is introduced along with its analysis protocol.

2.3.3.1 Attention control task

The field of psychology and psycholinguistics by extension has encountered difficulty in defining 'attention' accurately. Segalowitz (2009) claims we do not have one universally acknowledged definition of attention but cites Cohen, Aston-Jones, and Gilzenrat's (2004) views in which attention is described as "the emergent property of the cognitive system that allows it to successfully process some sources of information to the exclusion of others, in the service of achieving some goals to the exclusion of others" (91). The field of applied linguistics has also made efforts to define this intricate concept. For instance, Tomlin and Villa (1994) provide an overview of four different conceptions of attention. First, attention is understood as a limited-capacity system (i.e., a processing system that can handle only a small part of the information impacting on the individual); second, as a process of selecting critical information for further processing (selective attention); third, as a processing system that involves effort as opposed to less effortful automatic processing; and fourth, as a system involved in the control of information and action. Certainly, the relationship between language and attention (both attention-to-language and language-directed attention) has been widely studied in the past (for a number of samples see Rogers & Monsell, 1995; Segalowitz & Freed, 2004; Taube-Schiffnorman & Segalowitz, 2005). In this dissertation, an attention control task devised in Darcy, Park and Yang (2011) was implemented. Although this attention control task was originally intended to measure attention control in L2 English, in this dissertation it was used with the IM learners as well.

Given that the ability to rapidly shift attention to different level of linguistic information is not directly measured by working memory tests, and can be involved in selecting input for subsequent processing, a test was developed to measure attentional control in L2, the "Speeded Category Decision" task, which can be seen as an auditory version of the Wisconsin Card

Sorting Task (Heaton, 1981) or of the Dimensional Change Card Sort (Bialystok & Martin 2004).

Participants were presented with auditory stimuli consisting of an English word or non-word uttered by either a male or a female voice. Each of these four combinations was presented immediately following one of two possible questions: 'word?' or 'male voice?'. Participants had to answer the question for each stimulus by pressing on the keyboard a 'yes' button or a 'no' button. Figure 2.8 shows two sample questions as seen on the computer screen.

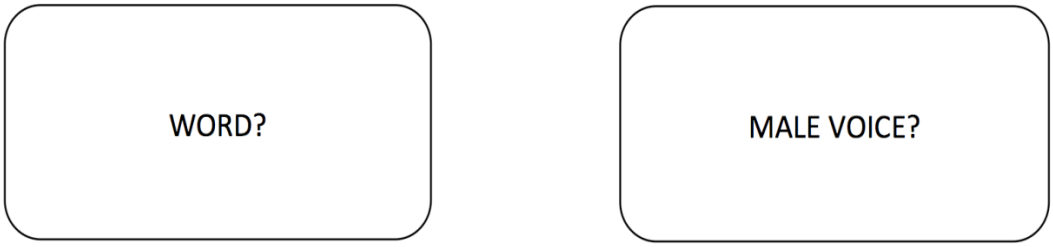

Figure 2.8. ACT only possible questions included in Darcy et al. (2011)

This task included ten practice trials followed by 100 randomized trials divided in four blocks of 25 trials each. Table 2.5 shows the different combinations that participants may have encountered in this task.

Table 2.5. Attention control task possible stimuli-question combinations with their answers

Stimuli Question	Female voice English word *'word?'*	yes	Male voice English word *'word?'*	yes
Stimuli Question	Female voice English word *'male voice?'*	no	Male voice English word *'male voice?'*	yes
Stimuli Question	Female voice non-word *'word?'*	no	Male voice non-word *'word?'*	no
Stimuli Question	Female voice non-word *'male voice?'*	no	Male voice non-word *'male voice?'*	yes

A correct answer to each question requires a different type of information to be extracted from the auditory stimulus: lexical vs. indexical. Depending on the sequence of questions (unpredictable), any given trial can be a shift trial (in which the participant needs to shift attention resources to extracting a different information from the signal), or a no-shift trial (in which the information to be focused on is the same as in the previous trial). A high ability to shift attention is seen in low error rates and fast RT in shift trials specifically.

All "no-shift" trials compose the baseline condition (CONDITION0). All "shift" trials (CONDITION1, 2, and 3) were further coded depending on the type of attentional shift involved in a correct answer to the trial question.

This distinction was important since different types of perseverations can be expected. On any shift trial, a participant may persevere with the previous answer regardless of the new question. This is called an "answer perseveration". This may yield an error (CONDITION1 and CONDITION3), but also conduct to a correct answer (CONDITION2). Table 2.6 provides an overview of the different combinations in all three conditions. A different type of perseveration concerns the level of information to extract from the signal. A participant with difficulties shifting attention from one level to the other may continue to answer the same question on the next trial, which we call a "level perseveration". Depending on the trial sequence, such a perseveration may lead to a correct answer (CONDITION3), or to an incorrect one (CONDITION1 and CONDITION2). There is also a fourth logical possibility of trial sequence (CONDITION4) where both types of perseveration lead to a correct answer. These trial sequences were avoided in the construction of the sequences, but occurred later through randomization (see below). They are not analyzed.

Table 2.6. Overview of all shift conditions in the Attention task; Conditions 1 and 3 will be compared to the baseline

SHIFT CONDITION			**1**				**2**				**3**	
MALE?	female	female	male	male	female	male	male	female	male	female	male	female
	word	nonword	word	nonword	nonword	word	nonword	word	word	word	nonword	nonword
correct answer	*no*	*no*	*yes*	*yes*	*no*	*yes*	*yes*	*no*	*yes*	*no*	*yes*	*no*
WORD?	word	word	nonword	nonword	word	nonword	nonword	word	word	nonword	word	nonword
	female	female	male	male	male	female	female	male	female	male	female	male
correct answer	*yes*	*yes*	*no*	*no*	*yes*	*no*	*no*	*yes*	*yes*	*no*	*yes*	*no*
Answer persev.	no	no	yes	yes	no	yes	y	no	yes	no	yes	no
(repeating answer)	wrong	wrong	wrong	wrong	wrong	wrong	wrong	wrong	right	right	right	right
Level persev.	no	no	yes	yes	yes	no	no	yes	no	yes	no	yes
(answering "male")	wrong	wrong	wrong	wrong	right	right	right	right	wrong	wrong	wrong	wrong

SHIFT CONDITION			**1**				**2**				**3**	
WORD?	word	nonword	nonword	word	word	nonword	word	nonword	word	nonword	word	nonword
	female	male	female	male	female	female	male	male	female	female	male	male
correct answer	*yes*	*no*	*no*	*yes*	*yes*	*no*	*yes*	*no*	*yes*	*no*	*yes*	*no*
MALE?	female	male	male	female	female	male	female	male	male	female	male	female
	word	nonword	nonword	word	nonword	word	nonword	word	nonword	word	nonword	word
correct answer	*no*	*yes*	*yes*	*no*	*no*	*yes*	*no*	*yes*	*yes*	*no*	*yes*	*no*
Answer persev.	yes	no	no	yes	yes	no	yes	no	yes	no	yes	no
(repeating answer)	wrong	wrong	wrong	wrong	wrong	wrong	wrong	wrong	right	right	right	right
Level persev.	yes	no	no	yes	no	yes	no	yes	no	yes	no	yes
(answering "word")	wrong	wrong	wrong	wrong	right	right	right	right	wrong	wrong	wrong	wrong

The major purpose of this task is to test the learner's ability to inhibit in particular the level perseverations. Therefore, the most important comparisons will be made between the baseline, that is the no-shift trials (CONDITION0), and shift trials drawn from conditions 1 or 3. The details of the analysis is explained in Chapter 4, section 4.5.

Darcy et al. (2011) used this task to address a possible correlation between attention control and phonological acquisition. The authors found that accuracy was very high for participants with an average of 84% correct responses. As expected, performance in the shift condition was significantly lower than in the no-shift condition ($p < .03$). Similarly, RT was slower in the shift condition compared to the no-shift condition ($p < .001$). In terms of individual differences in the attention shift ability, RT was negatively correlated with different measurements of phonological acquisition. In other words, participants with lower RT on the attention control task (reflecting a more efficient shift of attention among speech dimensions) obtained higher accuracy on two measures of successful suppression of L1 interference in L2 phonological processing (Darcy et al., 2012).

Although Darcy et al. (2011, 2012) used this task at one data collection time, in this dissertation learners performed this task twice, once at Time 1, and then at Time 3. The best conditions to help us define and describe attention control are 0, 1 (rate correct), and 3 (error rate). The challenge is to attribute an explanation in terms of the condition. Each participant receives an average reaction time for each condition, and RT differences could be used as an index for the level of control that learners possess towards specific properties of the input that they need to attend.

2.3.3.2 Semantic classification task

2.3.3.2.1 Spanish semantic classification task

The goal of the semantic classification task was to measure speed and accuracy of L2 lexical access. Speed of lexical access was operationalized in RT, measured in milliseconds. Accuracy was measured by considering the number of right and wrong answers. In this task, the stimuli were Spanish words. Modeled according to Segalowitz and Freed (2004), the current version consisted of a computerized test in which participants made speeded, two-alternative forced-choice animacy judgments (i.e., they had to decide if a word referred to a living or nonliving object) about single nouns presented on a computer screen (e.g., *el niño* 'the boy'= living; *el barco* 'the boat' = nonliving). For each stimulus, participants had to decide if each word represented a living or non-living entity as quickly as they could by pressing the '*animate*' or '*inanimate*' button on the keyboard. Figure 2.9 shows a sample word stimulus used in the Spanish version of the semantic classification task. Following Segalowitz and Freed (2004), nouns were presented with definite and indefinite articles to highlight the Spanish character of the words. Each version of the test began with ten warm-up trials followed by 100 experimental trials. This experimental version of the lexical access test yielded two measures, accuracy and speed of lexical access, as measured by RT.

At Time 1 participants were presented with a set of 100 new experimental trials (words). However, at Time 2, 50% of the words were from Time 1; the other 50% were new. Finally, at Time 3, the words were composed as follows: 25% from Time 1; 25% from Time 2; and 50% new. The stimulus words used in the IM and AH Spanish semantic classification task for Times 1, 2 and 3 are included in Appendix D.

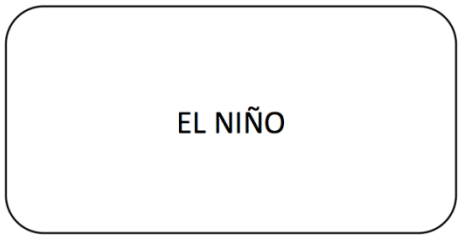

Figure 2.9. Spanish SCT sample stimulus item

The semantic classification task chosen for this study requires lexical access which entails word recognition (Camarazza & Brones, 1980; Segalowitz, Trofimovich, Gatbonton, & Sokovskaya, 2008). The theoretical principle behind this task is that in order to decide about the particular aspect of each stimulus, participants had to access each word in their mental lexicon to perform the task correctly. Animacy judgments have been used extensively in the recent literature (e.g., Phillips, Segalowitz, O'Brien, & Yamasaki, 2004; Segalowitz & de Almeida, 2002; Segalowitz & Frenkiel-Fishman, 2005; Segalowitz, Trofimovich, Gatbonton, & Sokolovskaya, 2008). In the field of psychology, animacy judgment tasks were created in reaction to the use of lexical decision tasks. According to Segalowitz (2009), lexical decision tasks require participants to perform "the somewhat unnatural task of distinguishing real words from non-words" (88). On the contrary, semantic classification tasks require animacy judgments in which participants focus on the meaning of the word that is presented to them.

2.3.3.2.2 English semantic classification task

The English version of the semantic classification task follows the same principles of the Spanish version previously described. The purpose of using this task was to replicate a portion of Segalowitz and Freed (2004) but in a different learning context and most importantly, while the participants were still in their overseas immersion setting. The stimulus words used in this task

were replicated from Segalowitz and Freed (2004).[12] Participants were presented with 100 English words in the experiment that included an indefinite and/or definite article. All stimuli were scrambled at each data collection point. Figure 2.10 shows a stimulus sample.

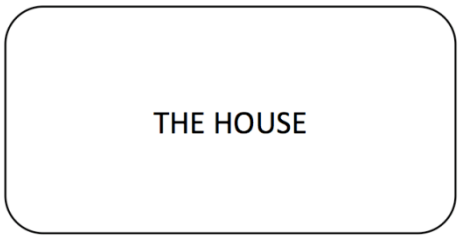

Figure 2.10. English SCT sample stimulus item

While the Spanish version of the semantic classification task had three different versions that were presented to the learners at three different time points, the English version was presented twice only to the participants (i.e., at the beginning and at the end of the program) in an effort to replicate the design obtained by Segalowitz and Freed (2004), by collecting the data *in situ* (instead of before and after the learners had lived in the abroad context). At both time points participants were presented with exactly the same version of the task. The stimulus words used in the IM and AH English semantic classification task for Times 1, 2 and 3 are included on Appendix D.

2.3.3.2.3 Analysis protocol of the Spanish and English semantic classification task

The Spanish and English semantic classification tasks were analyzed taking into account the following specifications: first, RT times above 2500 ms and below 300 ms were discarded. This resulted in a total of 5% of the data of the Spanish semantic classification task, and 2% of the data of the English semantic classification. Average accuracy and RTs were computed for

[12] The only tangible difference between the English semantic classification task in Segalowitz and Freed (2004) and the one included in the current dissertation is that participants were given 10 warm-up trials instead of 6.

each participant, and each condition and data were analyzed using the variables "animacy" and "old/new word".

2.3.3.3 Picture-naming task

The purpose of the picture-naming task was to measure speed and accuracy of L2 lexical retrieval. Speed was measured in milliseconds (i.e. RT) and corresponded to the time taken between the moment the picture appeared on the screen and the onset of the participant's response. Accuracy was measured in number of right and wrong responses (e.g., a word that did not correspond with the name(s) of the image that is shown). In this task the stimuli were composed by standardized black and white pencil drawings from the Snodgrass and Vanderwart (1980) database. Participants were asked to name the object image as soon as they could. In order to perform this task correctly, participants not only had to access the particular lexical item in their lexicon, but also retrieve it, and finally articulate the word. In this task, participants were presented with 100 different pictures at each test time.

To ensure baseline knowledge of the vocabulary items, a pilot study was carried out with 60 Midwest high school learners as well as with 40 college-level learners at a large Midwest institution. All participants were tested for knowledge on 453 pictures selected from the Snodgrass and Vanderwart corpus. Figure 2.8 shows a sample stimulus used in the picture-naming task. Participants were given a numbered sheet on which they had to write the Spanish word for the objects that they saw on a projected PowerPoint presentation.[13] The rationale behind this task was that an ideal picture-naming task should include words known by the participants in the study. The pilot results provided by these two groups of L2 learners helped to make a

[13] In this pilot study, each drawing was presented for three seconds, and participants were instructed to provide a Spanish word if they recognized the object or nothing when they did not know the name of the word. The task took approximately 45 minutes to complete. Although this task is substantially different from the picture-naming task, it provided a baseline to know what words the target populations of the current dissertation would knew, and therefore which ones should be included.

decision about what words would be included in the dissertation task. It was decided that the stimuli for the picture-naming task in this dissertation would be 100 words that at least 90% of the pilot learners knew.

Since the picture-naming task included oral production, all participants were tested individually in a quiet room. Like the other cognitive tasks included in the current dissertation, the picture-naming task was presented in DMDX (Forster and Forster, 2003). Participants sat in front of a computer and wore a Shure head-mounted microphone. Once the researcher was sure that participants were comfortable and ready they were instructed to follow the instructions pertaining to the picture-naming task on the screen before starting the task. Then, they performed a practice block to name ten pictures. Participants were presented with a fixation point first followed by a time-randomized picture in the center of the computer screen. Participants were instructed to respond by saying aloud the picture's name. Each picture was replaced with a fixation point after registering a response from the participant. An example of one of this picture stimulus is included on Figure 2.11.

Figure 2.11. Spanish PNT sample stimulus item from the Snodgrass and Vanderwart (1980) corpus

The software CheckVocal (Protopapas, 2006) was used to verify the RT measurements recorded by DMDX of each of the picture stimuli. Checkvocal provides sound player software accompanied by a spectrogram that allows the researcher to listen to each recording individually to confirm the appropriate RT measurement for each response. Although DMDX is able to provide RT measurements, CheckVocal allows further corroboration of each measurement as

well as to conveniently code them for accuracy by labeling responses as correct or incorrect. Following Sunderman and Kroll (2009), reaction times faster than 300ms or slower than 3000ms were removed from the analysis. Also, RTs that were 2.5 standard deviations above or below a participant's mean RT were excluded from the analysis and treated as outliers.

For the picture-naming task, I used DMDX (Forster & Forster, 2003) to run the script with a procedure similar to that of the semantic classification tasks, as well as Checkvocal (Protopapas, 2004), designed to manually access, analyze and label RT measurements involving speech. The analysis procedure involved the following steps. First, the DMDX script contained instructions to activate Recordvocal, a subcomponent of the DMDX software that records anything that comes through a microphone connected to the computer, from the onset of the stimulus, for as long as it is indicated in the script (3500ms). Once participants finished the picture-naming task at each time point, data were examined in spectrogram format using Checkvocal. With the help of this software, I was able to determine the RT for each individual stimulus uttered by each individual participant. In addition, Checkvocal allowed me to label the responses manually for right and wrong responses. I performed the analysis of each participant's picture-naming task recordings and acoustically determined the onset of each stimulus. Following Sunderman and Kroll (2009), RT measurements above 3000ms or below 300ms were discarded. This accounted for 6% of the data. For each participant and condition, average RT was the dependent variable. Independent variables were "animacy", "new/old word" and "number of syllable". The stimulus pictures used in picture-naming task for times 1, 2 and 3 are included on Appendix E.

2.4. IRB Approval

The Internal Review Board (IRB) of Indiana University approved the data collection protocol for this dissertation on March 25, 2010. The IM participants, who were minors, signed the informed consent form (along with their parents) at the IUHPFL orientation in Bloomington, Indiana, on April 10, 2010. All participants and their parents agreed to partake in the study after a brief explanation of the tasks and the time requirements of the experiments. In Fall 2010, AH participants were also presented with a similar informed consent document at NIU in DeKalb, Illinois, on August 24, 2010 (it was not necessary to have parental signature as they were all over 18 years of age), after a similar explanation of the tasks and time constraints was provided to them.

CHAPTER 3: ORAL FLUENCY ANALYSES FOR IM AND AH LEARNERS

3.1 Introduction

The goal of this chapter is twofold: first, to determine whether there are differences in proficiency scores for the AH and overseas IM groups; and second, to analyze how learners' oral fluency developed during their respective learning experiences in the US and overseas. In Chapter 3 fluency is operationalized through several measurements, each of them accounting for a very specific component of fluency. These measurements were chosen in an effort to obtain comprehensive rates for two aspects that have always been at the core of fluency research: speed of oral production and hesitation phenomena. The speed of speech delivery measures include: 1) total amount of speech spoken (measured in syllables), 2) total amount of spoken time (measured in seconds), and 3) rate of speech in syllables per second. The hesitation phenomena include 4) amount of filled pauses produced, 5) seconds per filled pause, and 6) syllables per filled pause. amount of production measured in syllables, amount of spoken time measured in seconds, development of oral language speed measured with rate of speech (i.e., measured in syllables per second, hesitation phenomena measured in filled pauses (i.e., raw scores), rate of seconds per filled pause and rate of syllables per filled pause.

The elicitation task chosen to elicit oral speech production for the IM and AH learners was a video retell that participants performed at three different points during their respective contexts. At each time, participants watched two different Tofield's Simon Cat's episodes for a total of six during their respective learning experiences. The results indicate that learners in the IM group were able to consistently develop a faster rate of speech across time, a higher rate of seconds per filled pauses, and a higher rate of syllables per filled pause, results that are commonly mirroring fluency improvement.

The AH group did also show an increase in the rate of seconds per filled pause, which at first sight, would also indicate some fluency development, however they did not increase their rate of speech; instead it remained steady across time. The data obtained through the transcription and analyses of the video retells were not only analyzed according to group but also according to low and high levels of proficiency, a group differentiation based on the proficiency scores obtained at Time 1 as well as individually. The discussion offers insight into issues relating to the development of second language oral fluency from a contextual perspective as well from the methodological challenge of incorporating new elicitation tasks and what that means for fluency research in general.

3.2 Analysis 1: Proficiency scores

This section presents the results of the proficiency test that was administered to the IM and AH learners and discusses their implications for the fluency tasks. Both groups of learners completed the same grammar test at the beginning of each learning context to guarantee that both groups were comparable in terms of grammar proficiency (see Chapter 2 for details). A series of Linear Mixed Models examined the effect of the independent variables: GROUP, TIME & PROFICIENCY on the dependent variable AVERAGE GRAMMAR SCORES.

3.2.1 Results

Initial inspection of the average score achieved by the AH and IM learners indicated that both groups had similar levels of proficiency at Time 1. Table 3.1 shows the average score for each group at Times 1 and 2 including the lowest and the highest score achieved by one of the learners in each specific group at each Time, as well as the standard deviation data for IM and AH groups.

Table 3.1. IM and AH proficiency scores at Times 1 and 2

	AVERAGE	*SD*	LOWEST SCORE	HIGHEST SCORE
IM T1	24.89	5.56	15	37
IM T2	32.11	5.67	20	40
AH T1	23.03	5.47	10	35
AH T2	24.76	4.97	18	40

There was no significant effect of GROUP at Time 1, $F(1, 32.40) = 1.861$, $p=.182$, which shows that there were no statistically significant differences between groups in terms of proficiency at the beginning of learning. At the end of the learning experience, however, the effect of group was highly significant, $F(1, 54) = 26.713$, $p<.001$; the IM learners were more accurate than the AH learners.

The effect of TIME on PROFICIENCY SCORES for the IM group alone was highly significant, $F(1,26) = 65.30$, $p<.001$. As for the AH group, a less pronounced effect of TIME was also found although still significant, $F(1,28) = 4.374$, $p=.046$. These results indicate that although both groups improved their grammar scores in a significant fashion over time, the IM group surpassed the AH one (as indicated by the significant effect of GROUP at time 2). It is important to highlight that although a statistical significance was found between Time 1 and Time 2 for the AH learners, the numeric difference between Times 1 and 2 was of 1.72 points only. There was a highly significant interaction of GROUP and TIME across time, $F(1,66) = 11.598$, $p<.001$, which indicates that the IM and AH groups did not behave in the same way at Times 1 and 2.

Given that the actual time that had passed between the test at T1 and T2 was different for both groups, and was much shorter for the IM group (six weeks) compared to the AH group (15 weeks), the larger improvement on this test could also be due that the IM group learners remembered the test better than the AH group learners. Although participants were never given the responses to the proficiency test at any time during their participation in the program, this

could be a possible caveat to this interpretation. Indeed, if the IM learners had remembered the test better, they may have performed better than the AH group because they were re-tested on it earlier than the AH learners.[14]

The data in Figure 3.1 represent two histograms with the distributional analysis of grammar proficiency results at Time 1 for each group (IM learners at the top, AH learners at the bottom). As can be seen, the results of both groups were so equally distributed that if the IM and AH histograms were to be superposed, they would match together almost exactly. In addition, both groups show a gap around the '24' score that separates very nicely low and high proficiency learners in both groups.

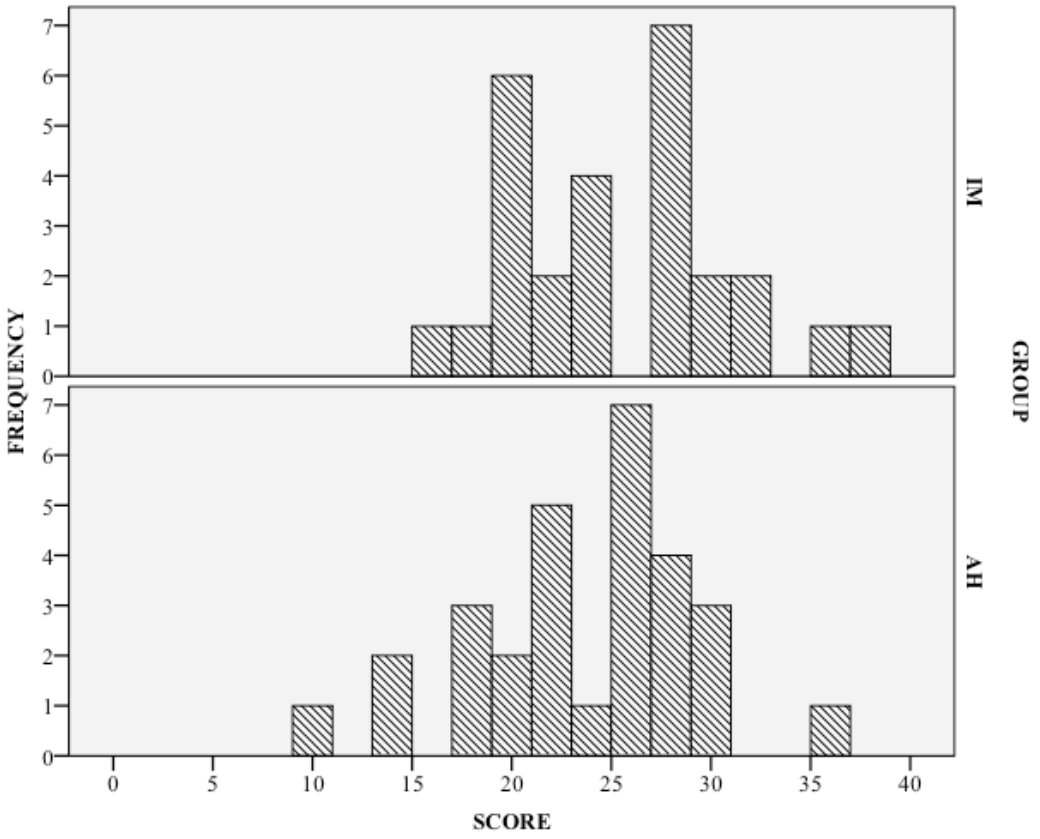

Figure 3.1. IM and AH histogram of proficiency scores at Time 1

[14] It would be important (in future work) to either statistically adjust for actual number of days between test and re-test, or to test another control group at the same time interval without any instruction, to see whether a shorter test-retest delay could have produced these results.

Figure 3.2 presents a distributional analysis of grammar scores for all IM and AH learners at Times 1 and 2. The solid boxes represent the IM and AH distributional proficiency scores at Time 1, and the checkered boxes represent the IM and AH distributional proficiency scores at Time 2. The figure shows that while both IM and AH learners had similar distributions at Time 1, the trend drastically changed at Time 2, when the IM learners received higher scores than AH one.

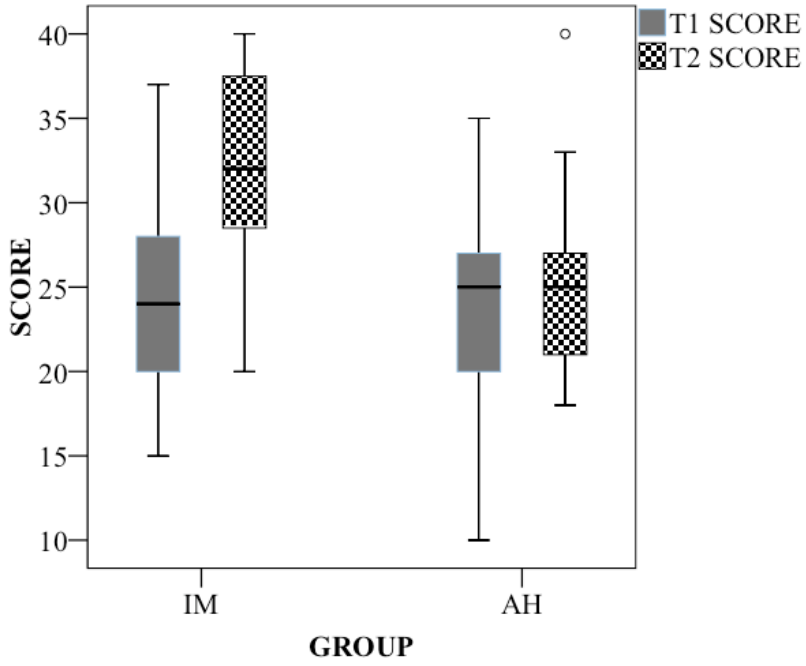

Figure 3.2. IM and AH proficiency score at Time 1 and at Time 2

After looking at the way the data were distributed, it was necessary to analyze the learner groups according to levels of proficiency. Separation of learners into separate proficiency groups was established based on their grammar score at Time 1. The descriptive analysis for IM and AH proficiency scores at Time 1 indicated that the median score for both groups was 24 points; this score was chosen as the threshold to separate the groups into high and low levels of proficiency at Time 1 and also at Time 2.

Table 3.2 summarizes the low and high proficiency scores for the IM and AH groups. Focusing on the IM group and more specifically, on the separation between low and high proficiency learners, it is worth underscoring that although low IM learners improve very little on average (i.e., their score increases from 20 points at Time 1 to 21 at Time 2), a Linear Mixed Model run to calculate the effect of TIME on SCORE for low proficiency IM learners indicated that there was a significant effect of TIME $F(1,13)=82.812$, $p<.001$. However, low proficiency AH learners had an average score of 19 at Time 1 and 21 points at Time 2, for which the Linear Mixed Model shows a significant effect for TIME on SCORE, $F(1,13) = 14.616$, $p=.002$. This indicates that overall low proficiency AH learners improved in terms of proficiency. The same applies to high proficiency AH learners, who received on average 27 points at Time 1 and 28 points at Time 2 in the proficiency test, a slight increase that was also statistically significant, $F(1,14)=14.616$, $p=.002$.

Table 3.2. Mean low and high proficiency scores for IM and AH groups

GROUP	Proficiency (Time 1)	Time 1	Time 2	Statistical analysis
IM (n=27)	Low (n = 14)	20 (SD = 2.50)	21(SD = 1.41)	F(1,13) = 82.812, p<.001
AH (n=29)	Low (n = 14)	19 (SD = 4.11)	21 (SD = 1.75)	F(1,13) = 14.616, p=.002
IM (n=27)	High (n = 13)	28 (SD = 3.45)	33 (SD = 4.86)	F(1,12) = 15.365, p=.002
AH (n=29)	High (n = 15)	27 (SD = 2.64)	28 (SD = 4.14)	F(1,14) =.332, p=.573

Figures 3.3 and 3.4 illustrate the distribution of low and high proficiency learners in the IM and AH groups at Times 1 and 2. In Figure 3.3 the two solid boxes represent the data extracted from the IM low and high proficiency scores at Time 1, and the two checkered boxes represent the data extracted from the IM low and high proficiency scores at Time 2. The proficiency data for both groups indicate that low and high proficiency IM learners improved their scores. In order to analyze the effect of TIME on SCORE two Linear Mixed Model analyses were conducted, one for each proficiency IM group, and the results revealed that the

effect for TIME was significant for high proficiency IM learners, F(1,12)=15.365, p= .002, as well as for low proficiency IM learners, F(1,13)=82.812, p<.001. The mean score was 20 points at Time 1 and 21 points at Time 2 for the low proficiency IM group, and 28 points at Time 1 and 33 points at Time 2 for the high proficiency IM learners. This change in mean value only corroborated the fact that 25 out of 27 IM participants received a higher score at Time 2. A close look at the IM data revealed that one learner received the same score at Time 1 and Time 2, 25 learners received higher scores at Time 2, and one learner received a lower score at Time 2.

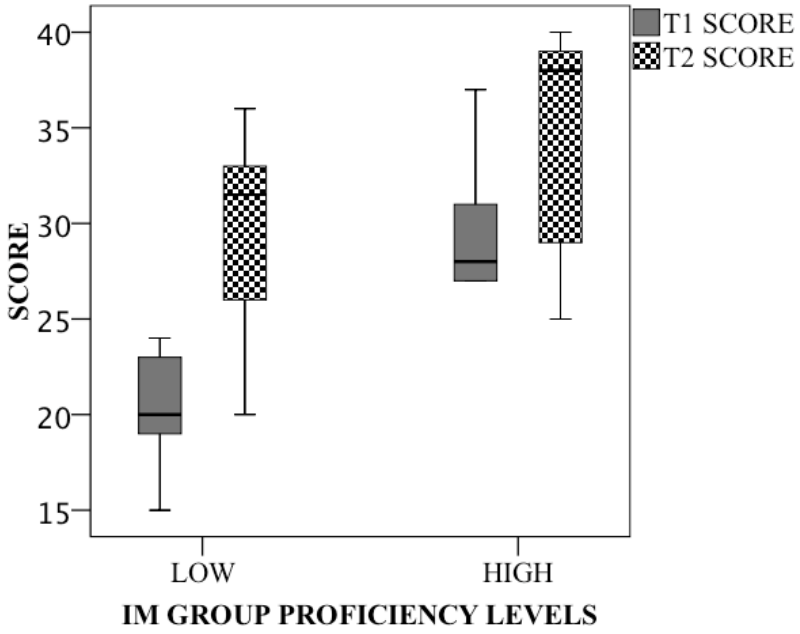

Figure 3.3. IM group proficiency levels at Times 1 and 2 [the value represented by the horizontal bar is the median]

In Figure 3.4 the solid boxes represent proficiency scores received by the low and high proficiency AH learners at Time 1, and the checkered boxes represent the proficiency scores received by the AH low and high proficiency learners at Time 2. For the AH group, both extremes of the boxes are very close, which indicates that there is little difference in the distributions. Closer inspection of the AH data revealed that one learner received the same score

at Times 1 and 2, 17 learners received higher scores at Time 2, and 11 received lower scores at Time 2.

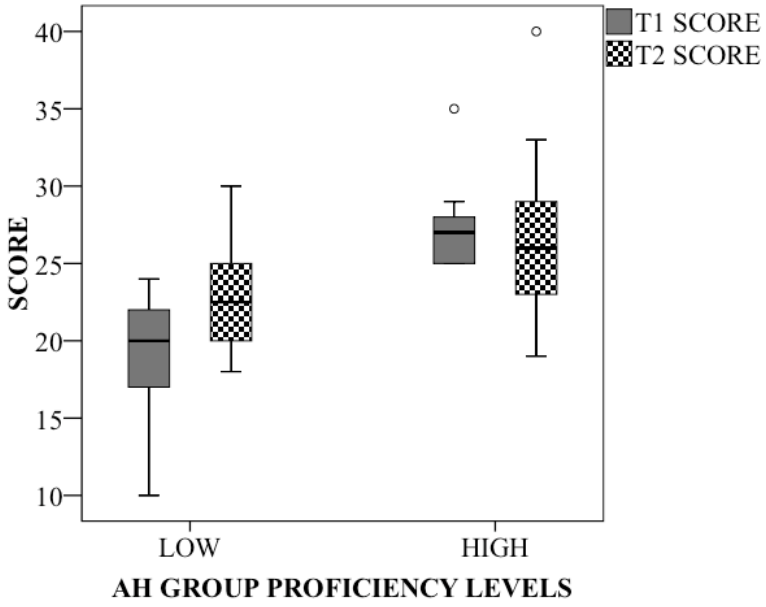

Figure 3.4. High and low AH proficiency levels at Times 1 and 2 [the value represented by the horizontal bar is the median]

3.2.2 Summary and discussion

The previous section described the difference in proficiency scores for IM and AH learners at the beginning and end of the respective learning experiences. The findings can be summarized as follows. There was a significant effect of TIME for both groups, but only at Time 2 were groups significantly different from each other, suggesting that the increase for the IM group was larger than for the AH group. This is visible in the Interaction between TIME and GROUP that was reported. Second, the data suggest a split into proficiency levels based on TIME and SCORE. Potential differences between proficiency groups were examined through Linear Mixed Models. It was found that there were significant differences between T1 and T2 for each subgroup (i.e., low proficiency AH, low proficiency IM, high proficiency AH and high proficiency IM).

The purpose of the proficiency test was to show that the two learner groups (IM vs. AH) were comparable in grammar skills at Time 1 but not at Time 2 and indeed, the results of the statistical analysis of the grammar proficiency scores showed that there were no statistical differences between the IM and AH groups at Time 1. In addition, the analysis showed that overall, both learner groups improved across time, although the AH learners did so to a lesser extent than the IM learners, as see through the significant effect of group at Time 2 only.

3.3 Analysis 2: Fluency measures

This section presents the analysis of fluency measures of L2 oral production.

3.3.1 Results: Fluency Measures for IM & AH Learners

This section introduces the quantitative analysis of the two video-retell tasks performed by the IM learners and AH learners at the beginning, middle and end of their learning experiences. The measures were described and motivated in chapter 2, section 2.3.2.1.

3.3.1.1 Analysis procedures

All 168 oral digital recordings (i.e., 56 speakers x 3 recording times = 168) were transcribed, and a special effort was put so that the orthography used in the transcriptions would reflect what the learners said. As was pointed out in Chapter 2, three measures contributed to quantification of speed of speech delivery (i.e., *amount of speech spoken in the L2*, *amount of time spoken in the L2*, and *rate of speech*) and three measures contributed to quantification of hesitation phenomena (i.e., *filled pauses, rate of seconds per filled pause, rate of syllables per filled pause*). Care was taken to identify and transcribe filled pauses and other hesitation phenomena. Then, each video retell was measured for amount of spoken time in seconds, and all words in every transcription were segmented into syllables.

Each individual *rate of speech* was calculated by dividing the *amount of speech spoken in the L2* in syllables for each video retell by the *amount of time spoken in the L2* in seconds taken by the learner to do the task. When rate of speech was calculated, filled pauses were left out, as they are not considered words in this dissertation (see Clark & Fox-Tree, 2001 for a different view). As for the other two rates included in the current analysis to account for Filled pauses, the *rate of seconds per filled pause* was calculated by dividing the total number of spoken time of each video retell by the total number of filled pauses produced in that specific fragment, and the *rate of syllables per filled pause* was similarly calculated with the difference of including syllables instead of seconds in the numerator.

As for the statistical analysis, a series of Linear Mixed Models were run to calculate the effect of TIME, GROUP and PROFICIENCY (i.e., low and high levels of proficiency) and the interactions of the three factors on SYLLABLES (i.e., *amount of spoken speech in the L2*), SECONDS (i.e., *amount of time spoken in the L2*), RATE OF SPEECH, FILLED PAUSES, RATE OF SECONDS PER FILLED PAUSE, and RATE OF SYLLABLES PER FILLED PAUSE. It is important to mention that when the IM and AH learners groups were separated according to low and high levels of proficiency, an additional analysis of the effect TIME and PROFICIENCY on each of the dependent variables was carried out for these learners subgroups.

3.3.1.2 Amount of speech spoken in the L2

In this section the results are presented separated per group. For IM and AH learners the data include an average score of each measurement used for the two videos combined. Table 3.3 shows the mean syllable production for videos 1 and 2 combined at each time point for IM and AH learners. It indicates that on average and for both videos combined, IM participants had a 35% increase in their syllable production between Time 1 and Time 2 and a 20% syllable

increase between Time 3 and Time 2 (a 49% increase between Time 1 and Time 3). A Linear Mixed Model analysis was run to analyze the effect of TIME on SYLLABLE PRODUCTION, and the results indicated that there was a significant effect of TIME between Time 1 and Time 2, $F(2,50)= 61.573$, $p<.001$; Time 1 and Time 3, $F(2,50)= 61.573$, $p<.001$; and Time 2 and Time 3, $F(2,50)= 61.573$, $p<.001$. Table 3.3 also shows the average syllable production for videos 1 and 2 combined across Time for the AH learners. Note that AH learners cannot break the overall mark of 300 syllables at any time point (IM learners produced an overall of 311 syllables for videos 1 and 2 combined at Time 1). There is, however, a 29% increase between Times 1 and 3, and there is a statistical significant effect of TIME on SYLLABLE PRODUCTION between Time 1 and Time 2, $F(2,54) = 15.068$, $p<.001$, Time 1 and Time 3, $F(2,54) = 15.068$, $p<.001$, but not between Time 2 and Time 3, $F(2, F(2,54) = 15.068$, $p=1.000$. Even with this increase, while IM learners doubled their overall syllable production (i.e., 50% increase between Times 1 and 3), AH learners developed a smaller increase (i.e., 29%).

Table 3.3. IM and AH average number of syllables produced for videos 1 and 2 combined

	Time 1	Time 2	Time 3
IM	311	493	604
AH	204	299	289

Another Linear Mixed Model was run to analyze the effect of GROUP on SYLLABLE PRODUCTION across time, and it turned out to be significant, $F(1,52) = 31.340$, $p<.001$. Also, when this effect was analyzed at each individual time, it was found that it was significant at Time 1, $F(1,52) = 17.795$, $p<.001$), Time 2 $F(1,52) = 15.417$, $p<.001$, but not at Time 3, $F(1,52) = 1.116$, $p=.296$. This indicates that although there is a significant effect of GROUP across time,

overall, the difference in syllable production between the IM and AH learners was more acute at the beginning than at the end of each of the contexts.

The data were also separated according to different levels of proficiency, and Figure 3.5 summarizes the average syllable production for IM learners with low and high levels of proficiency. A Linear Mixed Model was run to calculate the effect of TIME on SYLLABLE PRODUCTION for low proficiency learners, and the results indicated a highly significant effect of TIME between Time 1 and Time 2, $F(2,26) = 26.993$, $p=.001$; Time 1 and Time 3, $F(2,26) = 26.993$, $p<.001$; and Time 2 and Time 3 $F(2,26) = 26.993$, $p=.010$. A similar procedure was conducted for the high proficiency IM learners and the results indicated that for this subgroup of IM learners there was a highly significant effect for TIME on SYLLABLE PRODUCTION between Time 1 and Time 2, $F(2,24) = 34.860$, $p<.001$; Time 1 and Time 3, $F(2,24) = 34.860$, $p<.001$; and Time 2 and Time 3, $F(2,24) = 34.860$, $p=.011$. However, both groups did not differ from each other. Another global Linear Mixed Model was run to calculate the effect of PROFICIENCY on SYLLABLE PRODUCTION for the IM learner group and the results indicated that there was no significant effect of PROFICIENCY, $F(1,25) = 1.835$, $p=.188$. Then, another series of Linear Mixed Models was run to calculate the effect of PROFICIENCY at each Time point on SYLLABLE PRODUCTION, and no significant effect was found either at Time 1, $F(1,25) =1.975$, $p=.172$; at Time 2, $F(1, 25) = 1.617$, $p=.215$; or at Time 3, $F(1,25) = 1.487$, $p=.234$.

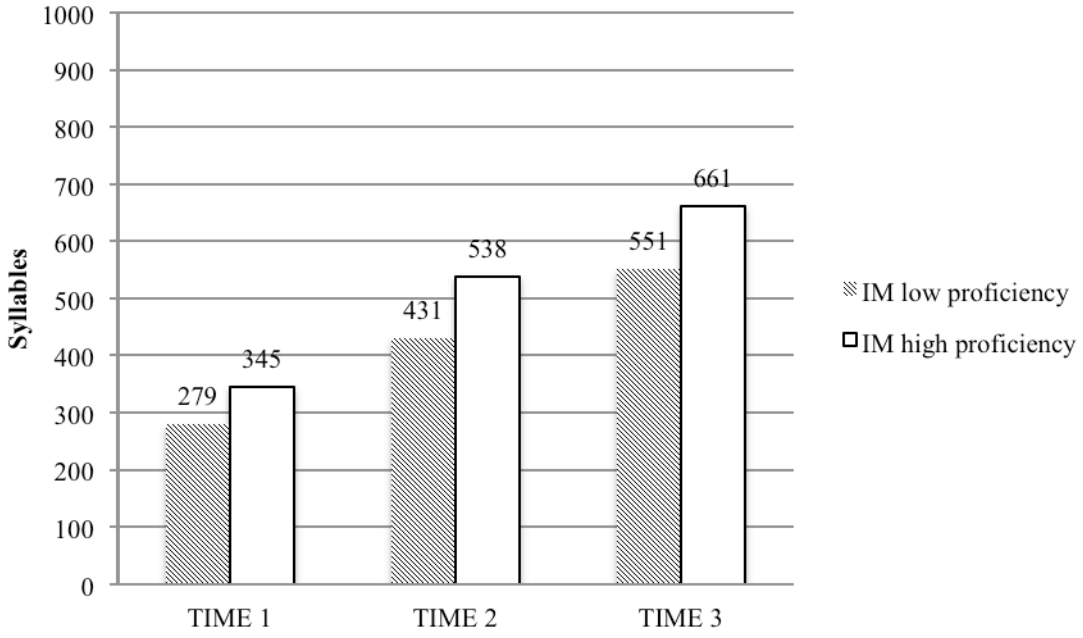

Figure 3.5. Average number of syllable produced by high and low proficiency IM learners across time

The AH data was also distributed according to low and high levels of proficiency, and a Linear Mixed Model was run to calculate the effect of TIME on SYLLABLE PRODUCTION for the low proficiency AH learners. The results revealed that there was a highly significant effect of TIME between Time 1 and Time 2, $F(2,26) = 12.74$, $p<.001$, Time 1 and Time 3, $F(2,26) = 12.74$, $p=.002$, but not between Time 2 and Time 3, $F(2,26) = 12.74$, $p=1.000$. As for the high proficiency AH learners the Linear Mixed Model indicated that the effect of TIME was not significant between Time 1 and Time 2, $F(2,28) = 3.886$, $p=.065$, Time 1 and Time 3, $F(2,28) = 3.886$, $p=.070$, or Time 2 and Time 3, $F(2,28) = 3.886$, $p=1.000$. Figure 3.6 summarizes the average syllable production for the AH group distributed for levels of proficiency. Just like for the IM learners, both proficiency subgroups did not differ either in the AH learners. A global Linear Mixed Model was run to analyze the effect of PROFICIENCY on SYLLABLE PRODUTION, and it was found that there were not statistical significant differences across time,

F(1,27) = .035, p=.853, or at each individual data collection time point: Time 1, F(1,27) = 1.412, p=.245; Time 2 F(1,27) = .459, p=.504, or Time 3, F(1,27) = .132, p=.719.

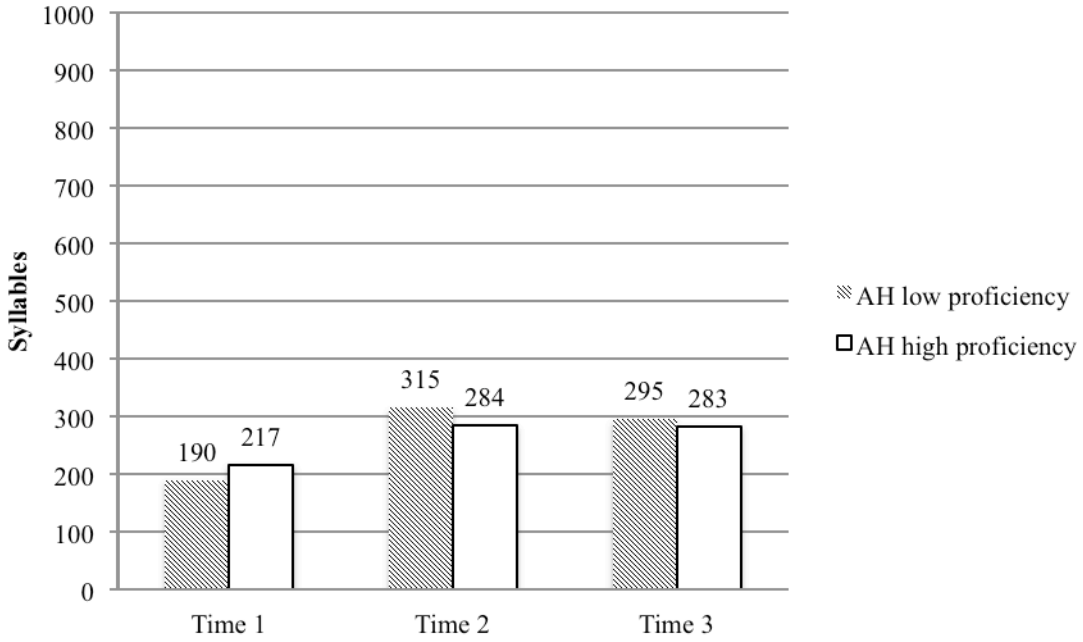

Figure 3.6. Average number of syllables produced by high and low proficiency AH learners across time

Also, interactions for TIME, GROUP and PROFICIENCY were run, and it was found that there was a significant interaction between TIME x GROUP on SYLLABLE PRODUCTION, F(2,104) = 21.105, p<.001, which indicates the AH and IM learner groups behaved differently at Time 1 than they did at Time 2. It was also found that there was not a significant interaction between TIME x PROFICIENCY on SYLLABLE PRODUCTION, F(2,104) = .063, p=.939, which indicates that the proficiency levels of the AH and IM learner groups across time did not have an effect on SYLLABLE PRODUCTION. The interaction PROFICIENCY x GROUP did not have a significant effect on SYLLABLE PRODUCTION either, F(1,52) =1.869, p=.177. Finally, the interaction of TIME x PROFICIENCY x GROUP was not significant, F(1,104) = 1.337, p=.267.

3.3.1.3 Amount of time spoken in the L2

This section presents the results of the calculations on the length of speech produced by the participants to retell the videos at each time point. Table 3.4 show the average total of seconds taken to retell videos 1 and 2 combined for the IM and AH learner groups. A Linear Mixed Model was run to analyze the effect of GROUP on SECONDS and a significant effect was not found across time, $F(1,52) = 3.910$, $p=.053$. However, a significant effect was found at Time 1, $F(1,52) = 5.037$, $p=.029$, and at Time 3, $F(1,52) = 5.830$, $p=.019$, but not at Time 2, $F(1,52) = .482$, $p=.490$. This indicates that at Time 1 and Time 2 the two groups were behaving in a much more different way than at Time 3.

IM learners increased their spoken time for videos 1 and 2 combined consistently across time. A Linear Mixed Model was run to analyze the effect of TIME on SECONDS for videos 1 and 2 combined, and the results showed that there was a significant effect between Time 1 and Time 3, $F(2,50) = 18.955$, $p<.001$, and between Time 1 and Time 2, $F(2,50) = 18.955$, $p<.001$, but not between Time 2 and Time 3, $F(2,50)=18.955$, $p=.082$. As for the AH group, their spoken duration for the two videos combined increased from Time 1 to Time 2 and then decreased at Time 3. Another Linear Mixed Model was run to analyze the effect of TIME on the SECONDS taken by AH learners to retell videos 1 and 2, and the results indicated that there was a statistically significant effect of TIME between Time 1 and Time 2, $F(2,54) = 13.711$, $p<.001$, and between Time 1 and Time 3, $F(2,54) = 13.711$, $p<.001$, but not between Time 2 and Time 3, $F(2,54) = 13.711$, $p=.731$. Figure 3.8 shows the average amount of time used by the AH learners to retell the video at each time.

Table 3.4. IM and AH groups average length of speech produced for videos 1 and 2 combined (in seconds)

	Time 1	Time 2	Time 3
IM	211	261	291
AH	163	241	223

With regard to proficiency levels, Figure 3.7 shows the average spoken time in seconds for high and low proficiency IM learners. A Linear Mixed Model was conducted to analyze the effect of TIME on SECONDS for low proficiency IM learners, and it indicated that there was a highly significant effect for TIME between Time 1 and Time 3, $F(2,26) = 5.991$, $p=.007$, but not for Time 1 and Time 2, $F(2,26) = 5.991$, $p=.323$, or Time 2 and Time 3, $F(2,26) = 5.991$, $p=.253$. As for high proficiency IM learners, it was found that there was a highly significant effect for TIME between Time 1 and Time 2, $F(2,24) = 15.104$, $p=.002$, Time 1 and Time 3, $F(2,24)=15.104$, $p<.001$, but not for Time 2 and Time 3, $F(2,24)=15.104$, $p=.504$. These differences indicate that although low and high proficiency learners increased their production in a significant fashion across time, high proficiency IM learners only showed a significant increase during the first three weeks of the program. Another Linear Mixed Model was run to analyze the effect of PROFICIENCY on SECONDS (i.e., the combined duration of the retelling of videos 1 and 2 by IM low and high proficiency learners), and the results indicated that there was not a significant across time, $F(1,77) = .288$, $p=.593$, or at individual times: Time 1, $F(1,25) = 1.255$, $p=.273$; Time 2, $F(1,25) = .014$, $p=.908$; nor Time 3, $F(1, 25) = .019$, $p=.892$.

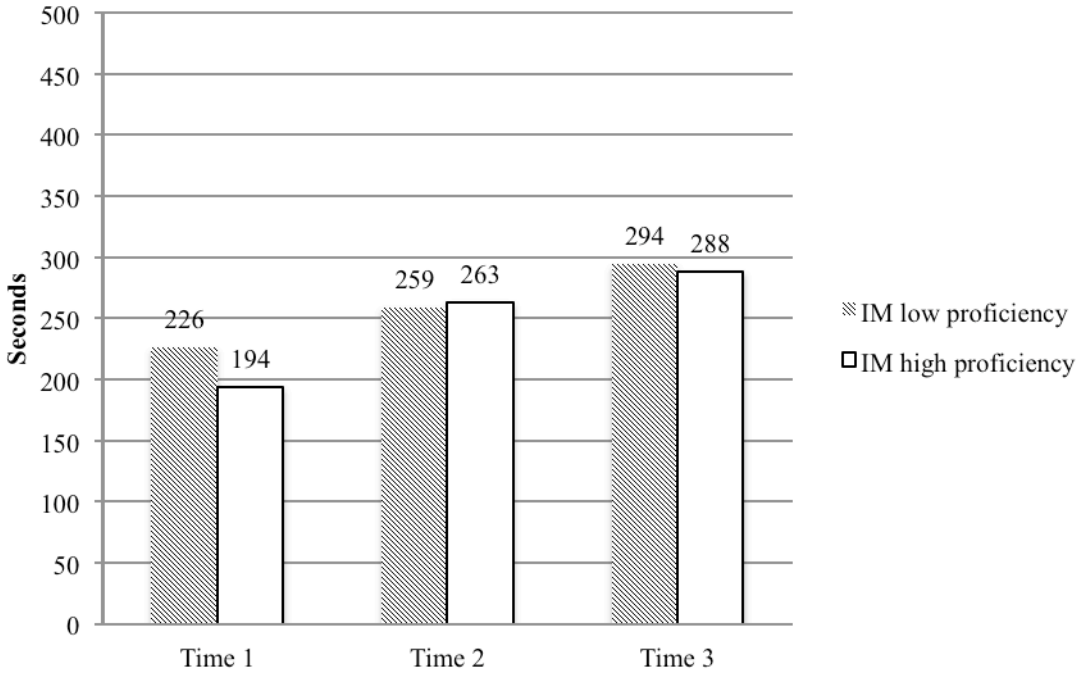

Figure 3.7. Average spoken time for low and high proficiency IM learners across time

As for AH learners Figure 3.8 shows the amount of time that low and high proficiency AH learners used to retell the two videos at each time point. A Linear Mixed Model indicated that there was a significant effect of TIME for low proficiency IM learners between Time 1 and Time 2, $F(2,26) = 13.471$, $p<.001$, Time 1 and Time 3, $F(2,26) = 13.471$, $p=.008$, but not between Time 2 and Time 3, $F(2,26) = 13.471$, $p=.243$. This indicates that overall low proficiency AH learners spoke for significantly more time during the first 7 weeks of classes as well as from the beginning to the end of their learning experience. However, when the same linear model was run to calculate the effect of TIME on SECONDS for high proficiency AH learners only, the results indicated that there was no significant effect between Time 1 and Time 2, $F(2,28) = 2,659$, $p=.201$, Time 1 and Time 3, $F(2,28) = 2,659$, $p=.141$, or Time 2 and Time 3, $F(2,28) = 2,659$, $p=1.000$. Thus, low proficiency AH learners increased the amount of speech produced across time, while high proficiency AH learners did not. Another Linear Mixed Model

was conducted to analyze the effect of PROFICIENCY on the SECONDS taken by AH learners to retell both videos and the results indicated the there was not statistically significant effect for proficiency across time, F(1,27) = 1.766, *p*=.194, or either at Time 1, F(1,27) = .060, p=.808, or Time 3, F(1,27) = .925, p=.345. The only significant effect of PROFICIENCY on SECONDS appeared at Time 2, F(1,27) = 4.246, p=.049,

Also, the effect of the interactions of the individual variables, GROUP, PROFICIENCY and TIME on SECONDS was calculated. The results of the statistical analysis indicated that there were no significant effects for either TIME x GROUP, F(2,104) = 2.775, p=.067; TIME x PROFICIENCY, F(2,104) = .504, p=.606; or PROFICIENCY x GROUP, F(1,52) =.458, p=.502. The only interaction that was found significant was TIME x PROFICIENCY x GROUP, F(2,104) = 3.447, p=.036, indicating that not all proficiency levels between groups behaved similarly over time.

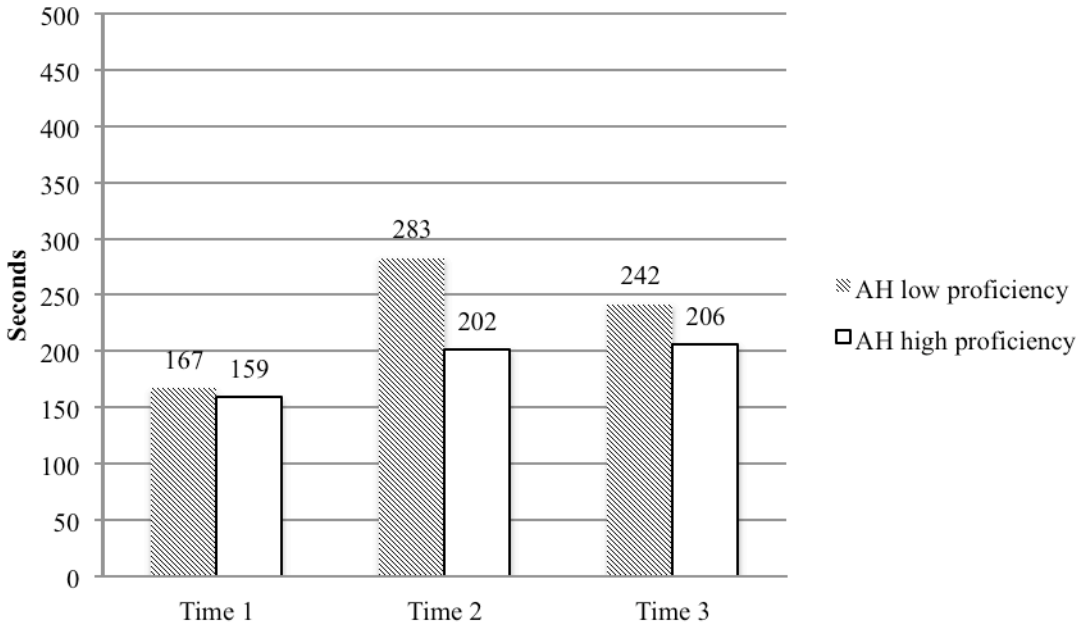

Figure 3.8. Average spoken time for high and low proficiency AH learners across time

3.3.1.4 Rate of speech

In this section the measurements of rate of speech obtained for both groups are presented. Table 3.5 presents IM and AH learner groups' average rates of speech for videos 1 and 2 combined. When the effect of GROUP on RATE OF SPEECH was analyzed it was found that there was a highly significant effect across time, $F(1,52) = 17.718$, $p<.001$, but not at Time 1, $F(1,52) = 2.088$, $p=.155$, which implies that the groups were not statistically different at Time 1, an aspect that changed at Time 2, $F(1,52) = 19.650$, $p<.001$, and at Time 3, $F(1,52) = 27.502$, $p<.001$. Although all videos were different, IM learners increased their combined rate of speech at each data collection time point in the video retell tasks, and the Linear Mixed Model that was conducted to analyze the effect of TIME on RATE OF SPEECH revealed a significant effect between Time 1 and Time 2, $F(2,50) = 38.185$, $p<.001$, Time 1 and Time 3, $F(2,50) = 38.185$, $p<.001$, and Time 2 and Time 3, $F(2,50) = 38.185$, $p<.001$. These results show that IM learners steadily increased their rate of speech during their immersion experience. As for the AH group, Table 3.5 shows the rate of speech measured in syllables per second for videos 1 and 2 combined, which increased from beginning to end, reaching 1.44 syllables per second at Time 3, although it does not reach the minimum overall 1.54 syllables per second of IM learners at Time 1. The statistical analysis indicated that there was no statistical effect for TIME at any data collection time point for AH learners: Time 1 and Time 2, $F(2,54) = 1.497$, $p=1.000$; Time 1 and Time 3, $F(2,54) = 1.497$, $p=1.000$, or Time 2 and Time 3, $F(2,54) = 1.497$, $p=.233$.

Table 3.5. IM and AH groups rate of speech measured in syllables per second for videos 1 and 2 combined.

	Time 1	Time 2	Time 3
IM	1.54	1.87	2.11
AH	1.39	1.33	1.44

The next step in the analysis included examination of differences based on level of proficiency, and Figure 3.9 summarizes this information for low and high proficiency IM learners. A Linear Mixed Model was conducted to analyze the effect of TIME on RATE OF SPEECH for low proficiency IM learners and indicated that there was a highly significant effect for TIME between Time 1 and Time 2, $F(2,26) = 31.122$, $p<.001$, Time 1 and Time 3, $F(2,26) = 31.122$, $p<.001$, and Time 2 and Time 3, $F(2,26) = 31.122$, $p=.035$, which indicates that low proficiency IM learners increased their rate of speech between all recording times. As for high proficiency IM learners a similar analysis indicated that there was a highly significant effect for TIME between Time 1 and Time 3 only, $F(2,24) = 11.562$, $p<.001$, however there was not a significant effect between Time 1 and Time 2, $F(2,24) = 11.562$, $p=.075$, and Time 2 and Time 3, $F(2,24) = 11.562$, $p=.071$. According to these results both subgroups improve their rate of speech overall but the low proficiency group improves between every time point.

The statistical analysis also examined the effect of PROFICIENCY on RATE OF SPEECH for IM learners and revealed an overall significant $F(1,25)=6.105$, $p=.021$. However, when this effect was examined based on each data collection time point, the highest significant difference between low and high proficiency learners was seen at first at Time 1, $F(1,25) = 10.105$, $p=.004$, then at Time 3, where the difference was barely significant $F(1,25)=4.448$, $p=.045$, however a significant effect was not observed at Time 2, $F(1,25) = 2.883$, $p=.102$. These results indicate that the difference between low and high proficiency learners was observed in the proficiency test as well as in rate of speech.

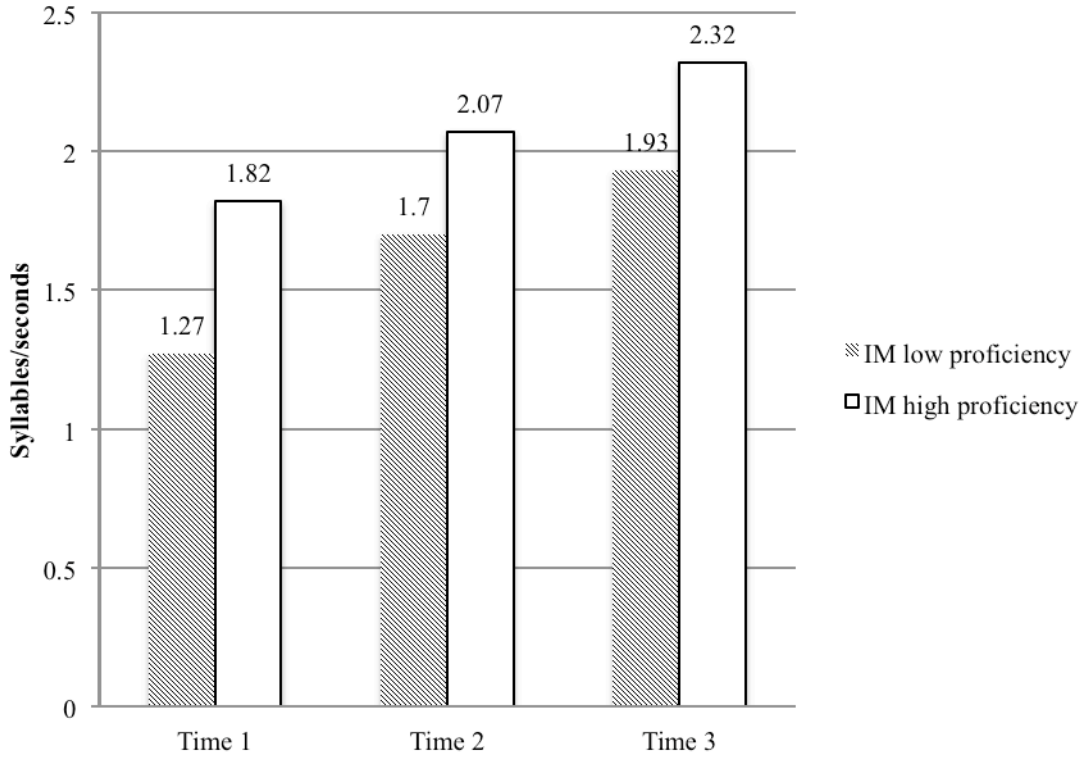

Figure 3.9. Average rate of speech in syllables per second for low and high proficiency IM learners

On the other hand, lack of significance was observed for both low and high proficiency AH learners once the data were separated according to these two levels of proficiency. Figure 3.10 shows the rate of speech for low and high AH learners. Indeed, for low proficiency AH learners there was no effect of TIME between Time 1 and Time 2, $F(2,26) = 1.729$, $p=1.000$, Time 1 and Time 3 $F(2,26) = 1.729$, $p=.346$, or Time 2 and Time 3, $F(2,26)=1.729$, $p=.371$, and the same applied to high proficiency AH learners: Time 1 and Time 2, $F(2,28) = 1.044$, $p=.481$; Time 1 and Time 3, $F(2,28) = 1.044$, $p=1.000$; Time 2 and Time 3, $F(2,28) = 1.044$, $p=1.000$. In the same fashion, there was no significant effect of PROFICIENCY on overall RATE OF SPEECH across time for AH learners, $F(1,27) = 2.478$, $p=.127$. However, when the effect of PROFICIENCY on RATE OF SPEECH was analyzed at each time point, AH learners showed a

significant difference in proficiency subgroups at Time 1, F(1,27) = 5.102, p=.032, but not at Time 2, F(1,27) = 2.262, p=.144 or Time 3, F(1,27) = .442, p=.512.

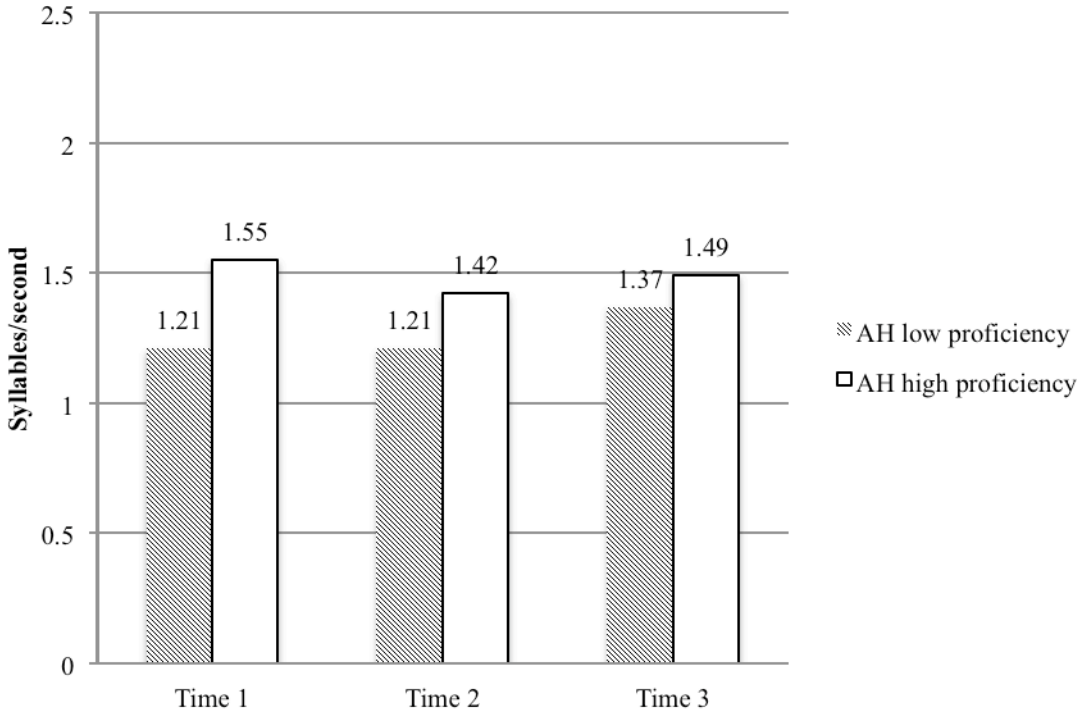

Figure 3.10. Average rate of speech for low and high proficiency AH learners across time

It is important to mention at this juncture that all the analyses have indicated that at Time 1 the independent variable GROUP had a significant effect on dependent variables such a SYLLABLES, F(1,52) = 31.340, p<.001, and SECONDS, F(1,52) = 3.910, p=.033. However, at Time 1 a significant effect for GROUP was not found on RATE OF SPEECH, F(1,53)=2.085, p=.155, an important finding that shows that prior to the beginning of their respective learning contexts, both groups were highly comparable, as they spoke with overall similar rates of speech, in addition to overall similar levels proficiency, as was shown in a previous section of this chapter.

A number of interactions were also analyzed for GROUP, PROFICIENCY and TIME on RATE OF SPEECH. The only interaction that was significant was TIME x GROUP, F(2, 104) =

17.721, p<.001, which indicates that IM and AH learners' RATE OF SPEECH did not behave in the same way across time. There were no significant interactions for either TIME x PROFICIENCY, F(2, 104) =2.321, p=.103; PROFICIENCY x GROUP, F(1, 52) =.906, p=.346; or TIME x PROFICIENCY x GROUP, F(2, 104) = .170, p=.844.

In addition to the analysis of independent variables such as GROUP, TIME, and PROFICIENCY on the dependent variable RATE OF SPEECH, it was decided to examine if there was an effect of VIDEO (i.e., the different pair of videos that participants watched at each data collection time point) on RATE OF SPEECH. Such follow-up analysis of rate of speech for each of the videos separately was also carried out (recall that to this point, the only results included in the current dissertation pertained to results combined for both video retells together at each time point).

Figure 3.11 shows IM learners' rate of speech for each pair of videos and shows that overall, they increased their rate of speech for each pair of videos across time. A Linear Mixed Model was conducted to analyze the effect of TIME on RATE OF SPEECH for the first video of each pair, and it showed a statistically significant effect for TIME between Time 1 and Time 2, F(2,52) = 21.609, p<.001, and Time 1 and Time 3, F(2,52) = 21.609, p<.001, but not for Time 2 and Time 3, F(2,52) = 21.609, p=.095. As for the second video of each pair, IM learners also showed a significant increase of rate of speech, and a Linear Mixed Model showed a significant effect of TIME on RATE OF SPEECH between Time 1 and Time 2, F(2,52) = 18.815, p=.037, Time 1 and Time 3, F(2,52) = 18.815, p<.001, and also between Time 1 and Time 2, F(2,52) = 18.815, p=.003.

When the difference between each video pair was studied at each time point, a significant effect on RATE OF SPEECH was found for Time 1 only, F(1,26) = 6.213, p=.019, and this

effect disappeared at Time 2, F(1,26) = .301, p=.588 and also at Time 3, F(1,26)=.000, p=.989, when the average rate of speech for IM learners was exactly the same for the two videos in the pair.

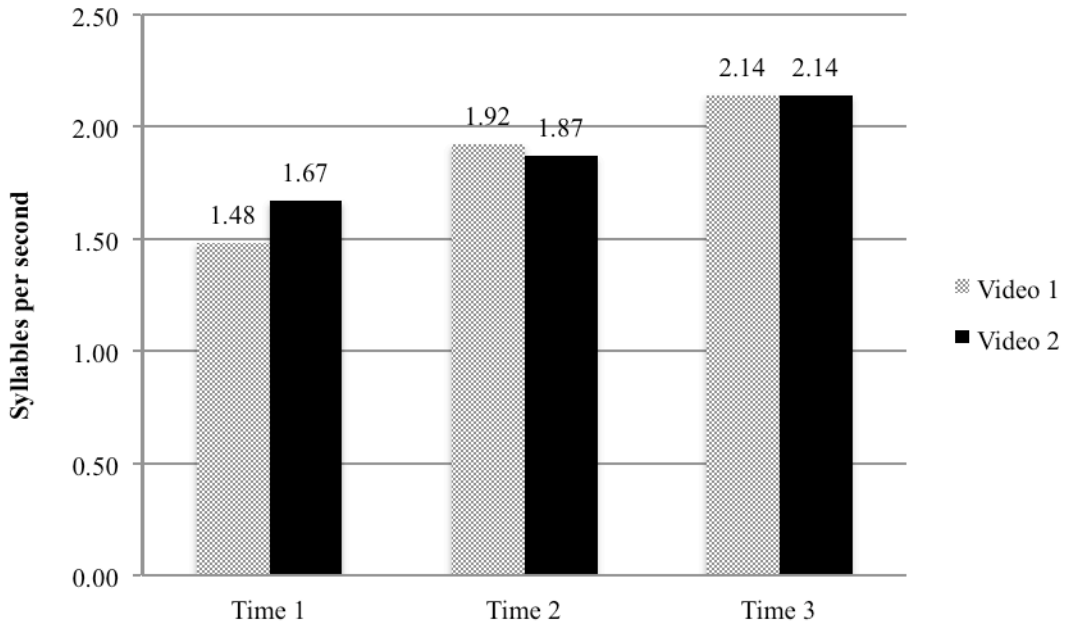

Figure 3.11. IM group mean rate of speech for videos 1 and 2 across time

In view of the significant effect of VIDEO at Time 1, the next step in the analysis was to examine the differences in each pair of videos according to levels of proficiency at each time point. The data are presented in Figure 3.12, and the statistical analysis revealed that for low proficiency IM learners there was not a significant effect of VIDEO on RATE OF SPEECH at any time: Time 1, F(1,13) = 1.515, p=.240, Time 2, F(1,13) = .092, p=.766, and Time 3, F(1,13) = 1.533, p=.238. However, when the same analysis was applied to high proficiency IM learners, a barely significant effect was found at Time 1, F(1,12) = 4.859, p=.048, an effect that was not found at Time 2, F(1,12) = .512, p=.488, or at Time 3, F(1,12) = .229, p=.641.

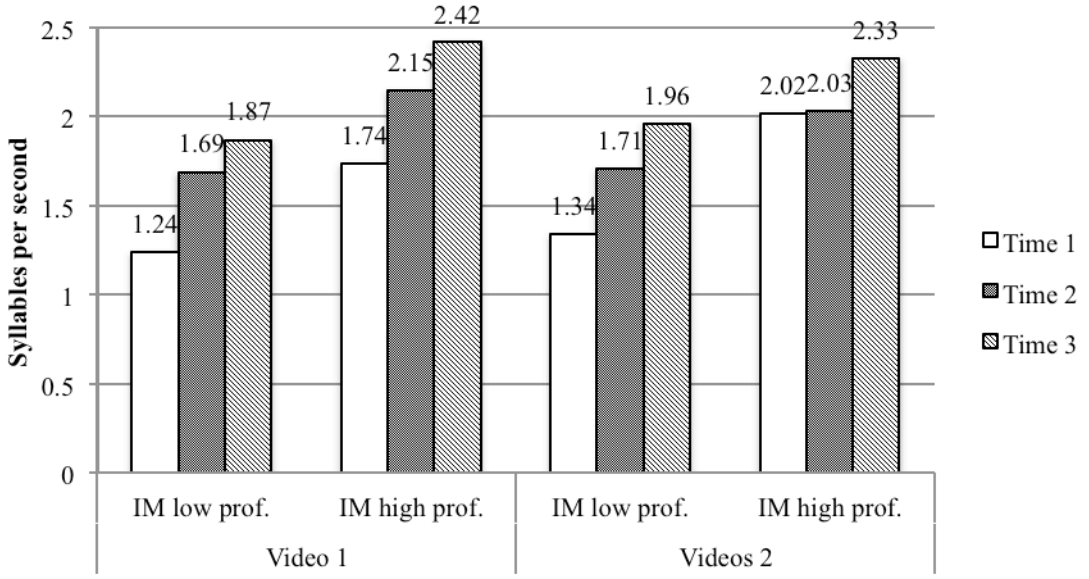

Figure 3.12. Average fluency scores for low and high proficiency IM learners

The same follow-up analysis was conducted with the AH learners, and Figure 3.13 shows the overall rate of speech for each video pair for the AH group at each time point. For this group the increase in rate of speech was not as drastic as it was for IM learners. Moreover, a series of Linear Mixed Models revealed that for the first video in each pair, TIME did not have a significant effect: Time 1 and Time 2, $F(2,56)=.523$, $p=.951$, Time 1 and Time 3, $F(2,56)=.523$, $p=1.000$, or Time 2 and Time 3, $F(2,56)=.523$, $p=1.000$. Similar statistical results were uncovered for the second video in each pair at Time 1 and Time 2, $F(2,56)=3.190$, $p=.739$, and Time 1 and Time 3, $F(2,56)=3.190$, $p=.525$, however, a small statistical significant effect was found between Times 2 and Time 3, $F(2,56)=3.190$, $p=043$, which indicates that AH learners retold the second video faster at Time 3 than they did at Time 2. Moreover, when the effect of VIDEO on RATE OF SPEECH was analyzed at each time point, it was found that the difference between the first and second videos at Time 3 was significant, $F(1,28) = 5.649$, $p=.025$, and that no significant results appeared at Time 1, $F(1,28) = .388$, $p=.539$ or Time 2, $F(1,28) = .087$, $p=.779$.

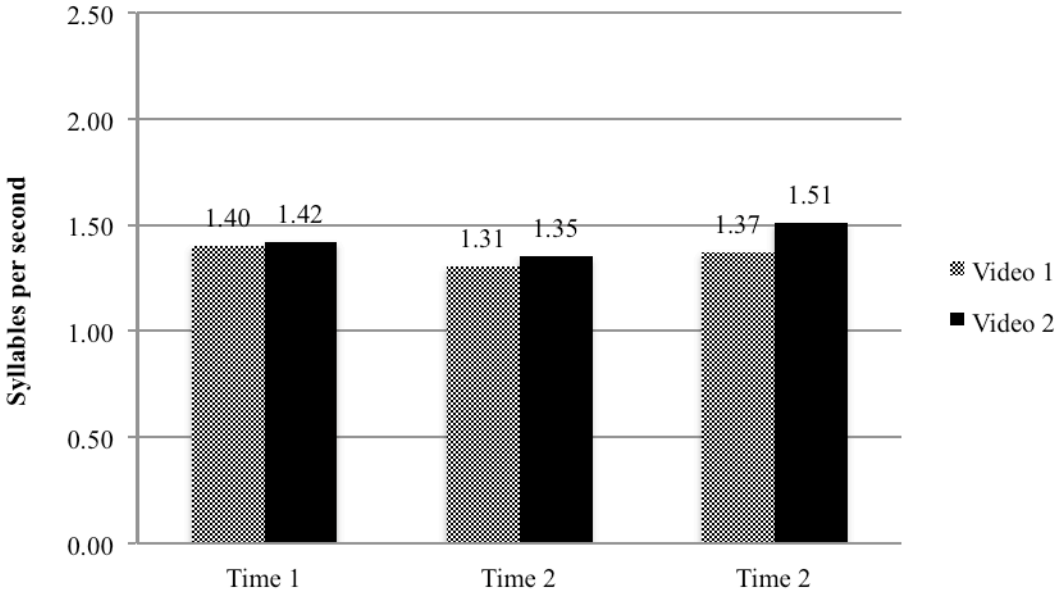

Figure 3.13. AH group mean rate of speech for videos 1 and 2 across time

The significant effect of VIDEO at Time 3 was analyzed taking into account the differences found according to levels of proficiency, and Figure 3.14 shows AH learners' rate of speech for each pair of videos distributed according to levels of proficiency. The statistical analysis did not reveal any significant effect of at any time neither for low proficiency AH learners: Time 1, $F(1,13) = 1.099$, $p=.313$, Time 2, $F(1,13) = 1.438$, $p=.252$, Time 3, $F(1,13) = 3.828$, $p=.273$, nor for high proficiency AH learners, $F(1,14) = 4.28$, $p=.524$, Time 2, $F(1,14) = 1.539$, $p=.235$, Time 3, $F(1,14) = 1.830$, $p=.198$.

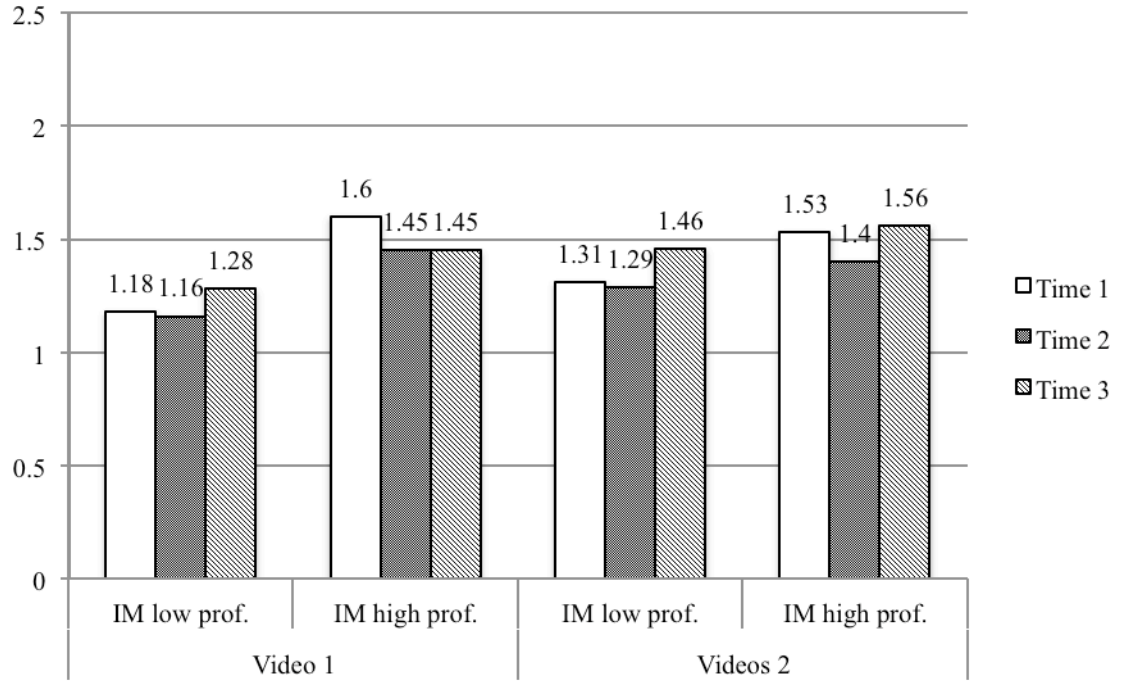

Figure 3.14. Average fluency scores for low and high proficiency AH learners

At this point, it is important to point out the difference between the IM and AH high and low proficiency levels and their differences with regard to rate of speech. A Linear Mixed Model analyzed the effect of LOW PROFICIENCY on RATE OF SPEECH for the IM and AH groups, and it was found that there was not a significant effect of LOW PROFICIENCY on RATE OF SPEECH for the IM and AH groups, $F(1,26)=.315$, $p=.579$, as well as there was not a significant effect of HIGH PROFICIENCY on RATE OF SPEECH for the IM and AH groups, $F(1,26)=1.77$, $p=.194$. This finding indicates that the previously mentioned significant effect of VIDEO at Time 3 responds to the overall difference in rate of speech for each video in the pair, a significant effect that does not apply when the groups are distributed according to different levels of proficiency.

3.3.1.5 Filled pauses

We now turn to the measures evaluating hesitation phenomena. This section presents the results of the Filled pause measurements obtained from the analysis of the video retell transcriptions of the IM and AH groups across three times in their respective learning contexts. Table 3.6 shows that the average filled pause production of IM and AH learners increases across time. The effect of GROUP on FILLED PAUSE PRODUCTION was analyzed across time, and it was found that there was a highly significant effect of GROUP, $F(1,52) = 19.958$, $p<.001$. This effect was also analyzed at each individual data collection time point, and it was found that it was also significant at Time 1, $F(1,52) = 9.754$, $p=.003$; at Time 2, $F(1,52) = 15.909$, $p<.001$, and at Time 3, $F(1,52) = 24.027$, $p<.001$. A Linear Mixed Model was conducted to compare the effect of TIME on FILLED PAUSE PRODUCTION on the IM learner group, and found a significant effect between Time 1 and Time 2, $F(2,50)=6.584$, $p=.013$, and Time 1 and Time 3, $F(2,50)=6.584$, $p=.006$, but not between Time 2 and Time 3 $F(2,50)=6.584$, $p=1.000$. As for the production of filled pauses by the AH learners, their average filled pause production across time remained very similar. Indeed, the statistical analysis did not yield any significant effect for TIME between Time 1 and Time 2, $F(2,54) = 1.925$, $p=.187$, Time 1 and Time 3, $F(2,54) = 1.925$, $p=.535$, or Time 2 and Time 3, $F(2,54) = 1.925$, $p=1.000$.

Table 3.6. Average number of filled pause produced for videos 1 and 2 combined.

	Time 1	Time 2	Time 3
IM	23	34	36
AH	9	11	10

These results were later separated according to levels of proficiency, and as Figure 3.15 shows, although low and high proficiency IM learners increased their average production of

filled pauses across time, low proficiency IM learners were more consistent in this increase. It is necessary to remind the reader, however, that these raw numbers are not ratios and do not take into account the amount of speech produced (either amount of spoken time or syllables produced). A Linear Mixed Model was run to determine the effect of PROFICIENCY on the average FILLED PAUSE production for the IM low and high proficiency learners, and the results indicated no significant effect of PROFICIENCY for the IM group neither across time $F(1,25)=.238$, $p=.630$, nor at each individual time point, Time 1, $F(1,25) = .078$, $p=.782$, Time 2, $F(1,25) = .071$, $p=.792$, Time 3, $F(1,25) = 1.838$, $p =.187$.

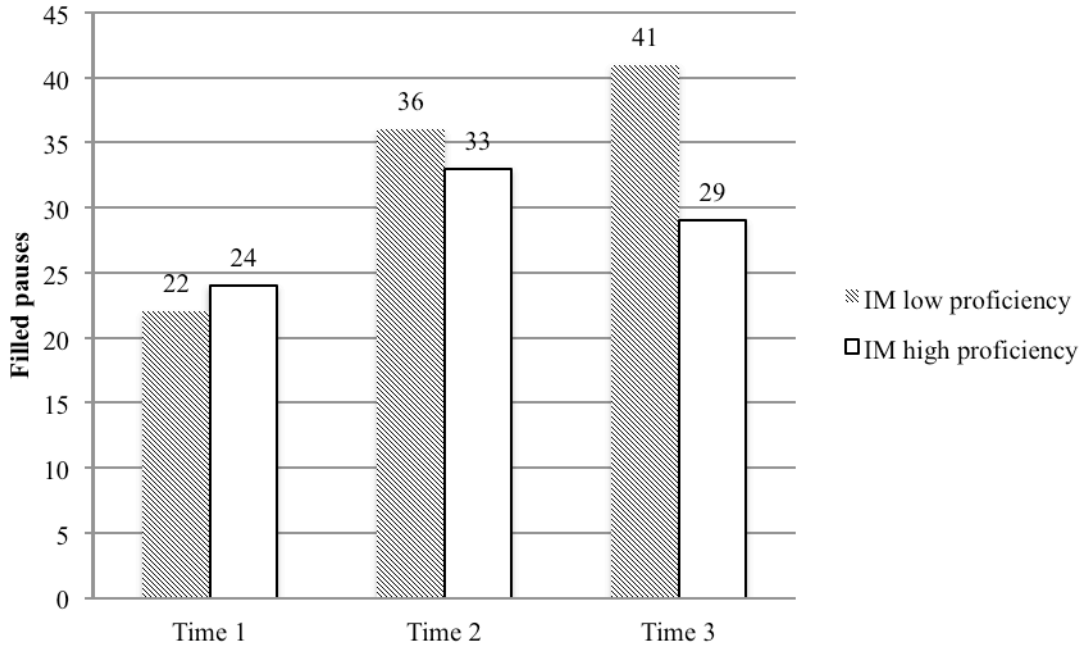

Figure 3.15. Average number of filled pauses produced by low and high proficiency IM learners across time

The average FP production for the AH group distributed according to low and high levels of proficiency appears in Figure 3.16. The statistical analysis indicated that there was not a statistically significant effect of PROFICIENCY on FILLED PAUSES for the AH learners across time, $F(1,27) = 2.362$, $p=.136$, nor at any particular time point: Time 1, $F(1,27) = .046$, $p=.832$, Time 2, $F(1,27) = .908$, $p=.349$, Time 3, $F(1,27) = .006$, $p=.938$.

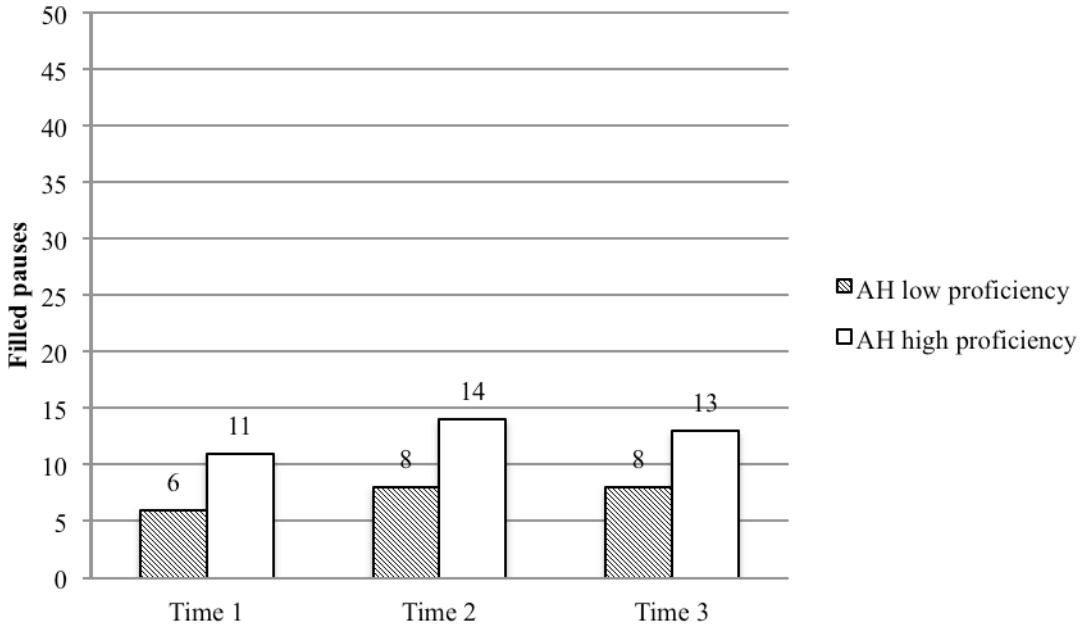

Figure 3.16. Average number of filled pauses produced by low and high proficiency AH learners across time

The effect of interactions for GROUP, PROFICIENCY and TIME on FILLED PAUSE PRODUCTION was also analyzed, and it was found that the interaction for TIME x GROUP was significant F(2, 104) = 3.795, p=.026. However other interactions were not: TIME x PROFICIENCY, F(2, 104) = 1.649, p=197; PROFICIENCY x TIME, F(1, 52) = .077, p=.783; TIME x PROFICIENCY x GROUP, F(2, 104) = 2.103, p=.127.

3.3.1.6 Seconds per filled pause

In order to analyze the role of filled pauses to a larger extent, an analysis of the rate of seconds per filled pause was also carried out. Table 3.7 shows that the average rate of seconds per filled pause for IM and AH learner groups. The effect of GROUP on RATE OF SECONDS PER FILLED PAUSE was also analyzed, and it was found that there was a significant effect of GROUP across time, F(1, 50) = 7.336, p=.009, as well as at Time 1, F(1,46) = 4.788, p=.034, but not at Time 2, F(1,48) = .136, p=.714, or at Time 3, F(1,47) = .139, p=.711. The IM learner group maintained very similar rates across time; although it seems to decrease slightly, the

statistical analysis showed no effect of TIME on RATE OF SECONDS PER FILLED PAUSE between Time 1 and Time 2, F(2,50) = .695, p=1.000, Time 1 and Time 3, F(2,50) = .695, p=.818, or Time 2 and Time 3 F(2,50) = .695, p=1.000. It is worthwhile to mention that the IM learners decreased their rate of seconds per filled pause across time during their abroad stay.

Table 3.7. Average rate of seconds per filled pause for videos 1 and 2 combined for IM and AH learner groups.

	Time 1	Time 2	Time 3
IM	18	15	14
AH	28	27	34

The role of PROFICIENCY was also analyzed at each time point, and Figure 3.17 shows the average rate of seconds per filled pause for low and high proficiency IM learners across time. A significant effect of PROFICIENCY was not found either across time, F(1,25) = 1.224, p=.279), or at any time point: Time 1, F(1,25) = .296, p=.591, Time 2, F(1,25)=2.529, p=.124, Time 3, F(1,25) = 2.175, p=.153. When the data were distributed for analysis according to high and low levels of proficiency, opposite patterns appeared for the two levels of proficiency, as high proficiency IM learners increased their rates from Time 1 to Time 2 (and remained stable at Time 3), and low proficiency IM learners decreased their rate of seconds per filled pause, or in other words, produced more filled pauses in less time. The statistical analysis showed that there was an effect of TIME on RATE OF SECONDS PER FILLED PAUSE between Time 1 and 2 for low proficiency learners, F(2,26) = 7.443, p=.014, between Time 1 and Time 3, F(2,26)=7.443, p=.005, but not between Time 2 and Time 3, F(2,26)=7.443, p=1.000. On the other hand, the balanced rate of seconds per filled pause for high proficiency IM learners did not show any significant effect of TIME on RATE OF SECONDS PER FILLED PAUSE between

Time 1 and Time 2, F(2,24) = .369, p=1.000, Time 1 and Time 3, F(2,24) = .369, p=1.000, or Time 2 and Time 3, F(2,24) = .369, p=1.000.

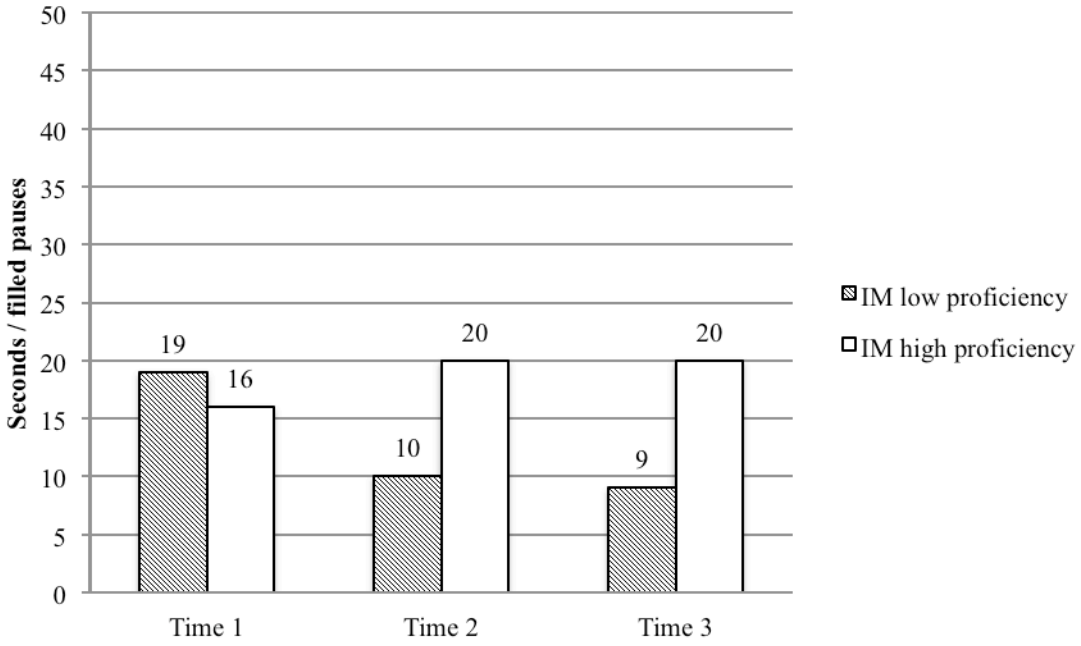

Figure 3.17. Average rate of seconds per filled pause for low and high proficiency IM learners across time

When the AH group was separated according to high and low levels of proficiency, it could be seen that on average, low proficiency learners slightly decreased their rate of seconds per filled pause between Times 1 and 3, which indicates more filled pauses in less time. High proficiency AH learners increased their rates of seconds per filled pause, or in other words spoke for more time without filled pauses. Figure 3.18 shows these results; the statistical analysis indicated that for low proficiency AH learners there was no effect for TIME at any time point (i.e., Time 1 and Time 2, F(2,25) = .469, p=1.000, Time 1 and Time 3, F(2,25) = .469, p=1.000, or Time 2 and Time 3, F(2,25) = .469, p=.631. The statistical analysis also indicated that there were no significant effects either for high proficiency learners between Time 1 and Time 2 (F(2,20) = .861, p=.724), Time 1 and Time 3 (F(2,20) = .861, p=.900), Time 2 and Time 3 (F(2,20) = .861, p=1.000).

As for the effect of PROFICIENCY on RATE OF SECONDS PER FILLED PAUSE for the AH learner group, no effect was found on rate of seconds per filled pause across time, F(1,25)=.013, p=.912, or at any individual time point, Times 1, F(1,23) = .007, p=.935, Time 2, F(1,21) = .871, p=.361, or Time 3, F(1,22) = .066, p=.799.

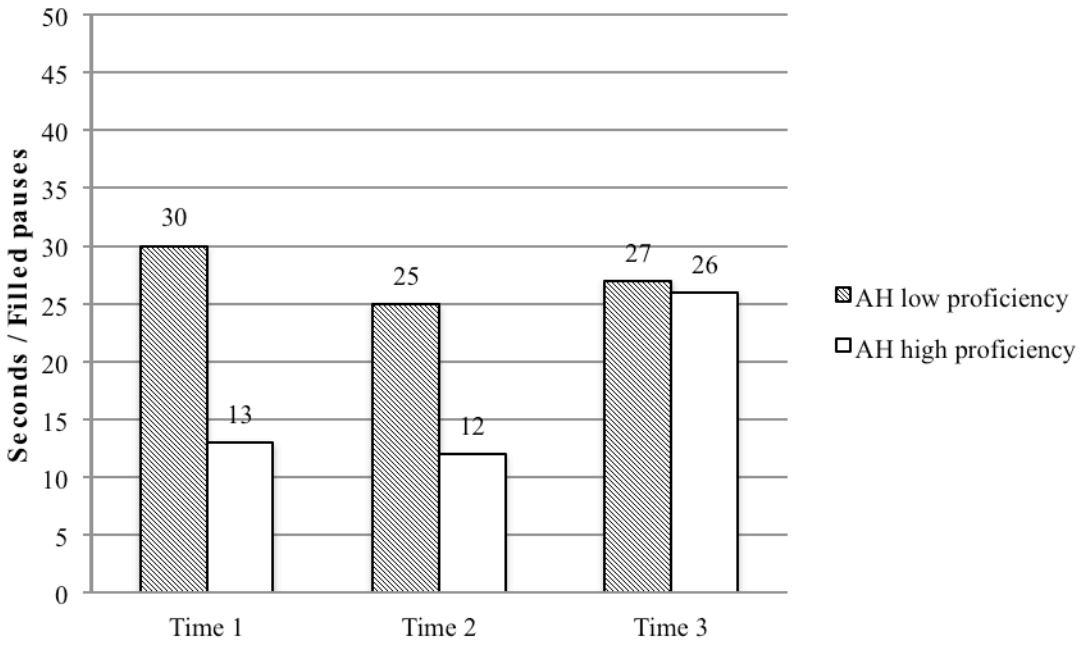

Figure 3.18. Average rate of seconds per filled pause for low and high proficiency AH learners across time

A number of interactions were calculated for GROUP, TIME and PROFICIENCY, and it was found the none of the interactions were significant: TIME x GROUP, F(2, 95) = .821, p=.443; TIME x PROFICIENCY, F(2, 95) = 1.006, p=.369; PROFICIENCY x GROUP, F(1, 50) = .232, p=.632; TIME x PROFICIENCY x GROUP, F(2, 95) = .860, p=.427.

A follow-up analysis indicated that an effect of VIDEO on rate of seconds per filled pause was not found for IM learners across time F(1,132) = .004, p=.947, or at any individual time point: at Time 1, F(1,26) = .468, p=.500, at Time 2, F(1,26) = .145, p=.707, and at Time 3, F(1,26) = 1.804, p=.191. AH learners did not have striking overall differences across time. A Linear Mixed Model was run to analyze the effect of TIME of RATE OF SECONDS PER

FILLED PAUSE, and the results indicated that there was no effect of TIME between Time 1 and Time 2, F(2, 45) = .468, p=1.000, Time 2 and Time 3, F(2, 45) = .468, p=1.000, or Time 2 and Time 3, F(2, 45) = .468, p=1.000. Continuing with the follow-up analysis carried out of the effect of VIDEO on RATE OF SPEECH, the effect of VIDEO on RATE OF SECONDS PER FILLED PAUSE was also analyzed. For AH learners such effect was not found across time, F(1, 142) = 1.425, p=.235, or at any particular time point: at Time 1, F(1, 28) =.411, p=.527; at Time 2, F(1,28) = .850, p=.364; at Time 3, F(1, 28) = 2.108, p=.158.

3.3.1.7 Syllables per filled pause

Since the results based on the rate of seconds per filled pause for the IM learners showed that the learners in the video retell task were producing filled pauses at a very different rate than what had been reported for similar learners (in identical learning contexts) (García-Amaya, 2010; García-Amaya & Filgueras-Gómez, 2011), a decision was made to explore another rate related to filled pauses, that is, the rate of syllables per filled pause. Instead of focusing on TIME, this rate accounts for the intervals of syllable production between which a filled pause is produced. Since the amount of speech and the spoken time were related to both groups of learners (more on correlations on chapter 5), the results of the rate of syllables per filled pause were expected to be along the same lines. Table 3.8 presents the average rate of syllables per filled pause for IM and AH learner groups. The effect of GROUP on RATE OF SYLLABLES PER FILLED PAUSE was also analyzed, and it was found that it was not statistically significant across time, F(1,49) = 2.160, p=.148, nor at each individual data collection time point: at Time 1, F(1,48) = 3.196, p=.080; at Time 2, F(1,46) = .829, p=.367; or at Time 3, F(1,47) = .982, p=.327. First, it shows that the overall tendency of the IM group is to slightly increase rate of syllables per filled pause, or in other words, to produce vaguely bigger chunks of speech per filled pause. However, when

the effect of TIME on the RATE OF SYLLABLES PER FILLED PAUSE was investigated, the analysis revealed that there was not a significant effect of TIME between Times 1 and 2, F(2,50) = .594, p=1.000, Times 1 and 3, F(2,50) = .594, p= .876, or Times 2 and 3, F(2,50) = .594, p=1.000. For this rate, a follow-up analysis was also conducted to analyze the effect of VIDEO on RATE OF SYLLABLES PER FILLED PAUSE for IM learners, which did not yield any significant results, neither across time, F(1,132) = .120, p=.730, nor at any given time point: Time 1, F(1,26) = .044, p=.835, Time 2, F(1,26) =.482, p=.494, Time 3, F(1,26) = 1.946, p=.175.

Table 3.8. Average rate of syllables per filled pause for IM and AH learner groups

	Time 1	Time 2	Time 3
IM	26	28	32
AH	39	35	51

As for the AH learners, overall, they decreased their rate of syllables per filled pause at Time 2 and then increased again at Time 3. Also, there was not a significant effect of TIME ON RATE of SYLLABLES PER FILLED PAUSE between Time 1 and Time 2, F(2,47) = .773, p=.1.000, Time 1 and Time 3, F(2,47) = .773, p=.992, or Time 2 and Time 3, F(2,47) = .773, p=.769. In addition to this analysis, the effect of VIDEO was examined, and n statistical significant effect was not found across time, F(1,142) = 1.356, p=.246, nor at any specific time point: Time 1, F(1,28) = .394, p=.536, Time 2, F(1,28) = 1.353, p=.255, Time 3, F(1,28) = 1.223, p=.278.

An additional analysis was carried out to study the role of proficiency on rate of syllables per filled pause, and Figure 3.19 shows the rates for high and low proficiency IM learners. According to this distribution, low proficiency and high proficiency IM learners started with very

similar rates at Time 1, but then, at Times 2 and 3 their rates of syllables per filled pause became more dissimilar. The statistical analysis revealed that the effect of TIME on the RATE OF SYLLABLES PER FILLED PAUSE did not yield significant results for either subgroup (i.e., for low proficiency IM learners, Time 1 vs. Time 2, $F(2,26) = 1.873$, $p=.414$, Time 1 vs. Time 3, $F(2,26) = 1.873$, $p=.255$, Time 2 vs. Time 3, $F(2,26)=1.873$, $p=.1.000$; and for high proficiency IM learners, Time 1 vs. Time 2, $F(2,24) = 1.399$, $p=1.000$, Time 1 vs. Time 3, $F(2,24) = 1.399$, $p=.344$, Time 2 vs. Time 3, $F(2,24) = 1.399$, $p=1.000$. Probably, the small sample size of each proficiency subgroup within the IM group caused the Linear Mixed Model to have problems with the iterations, and thus, it did not indicate any significant effect for PROFICIENCY on RATE OF SYLLABLES PER FILLED PAUSE across time $F(1,25) = 2.915$, $p=.100$, nor at any individual time point (i.e., Time 1, $F(1,25) = .466$, $p=.501$; at Time 2, $F(1,25) = 3.095$, $p=.091$; or at Time 3, $F(1,25) = 3.3037$, $p=.094$), although at Times 2 and 3 the difference between groups does approach significance.

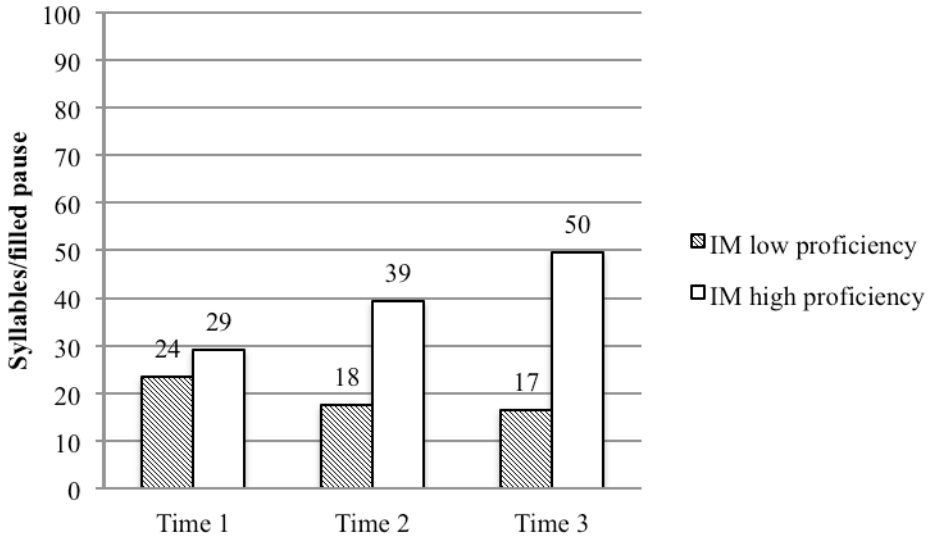

Figure 3.19. Average rate of syllables per filled pause for low and high IM learners across time

The next step in the analysis of the rate of syllables per filled pause for the AH learners was to analyze the distribution of results according to levels of proficiency. Figure 3.20 shows that while the rate remains very similar across time for low proficiency AH learners, who increase the amount of seconds per filled pause very slightly, it fluctuates more for high proficiency learners. The statistical analysis showed no statistical effect of TIME neither for low proficiency AH learners (i.e., Time 1 vs. Time 2, $F(2,19) = .387$, $p=1.000$, Time 1 vs. Time 3, $F(2,19) = .387$, $p=1.000$, Time 2 vs. Time 3, $F(2,19) = .387$, $p=1.000$), nor for high proficiency AH learners (i.e., Time 1 vs. Time 2 $F(2,25) = 24.571$, $p=1.000$, Time 1 vs. Time 3, $F(2,25) = .834$, $p=1.000$, Time 2 vs. Time 3 $F(2,25) = .834$, $p=.650$). Moreover, when the effect of PROFICIENCY on SYLLABLES PER FILLED PAUSE was analyzed statistically, it showed no significant effect across time (i.e., $F(1,24) = .231$, $p=.635$), or at each individual time point (i.e., at Time 1, $F(1,23) = .419$, $p=.524$, at Time 2, $F(1,21) = 3.170$, $p=.089$, or at Time 3, ($F(1,22) = 0.25$, $p=.877$).

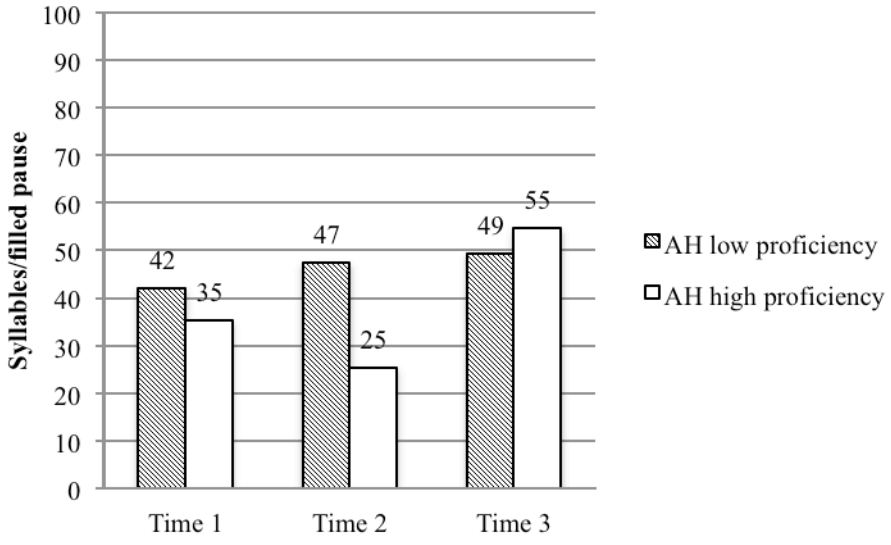

Figure 3.20. Average rate of syllables per filled pauses for low and high AH learners across time

Also, the effects of the interactions including GROUP, PROFICIENCY and TIME on RATE OF SYLLABLES PER FILLED PAUSE were calculated. It was found that there were no significant effects for any of the interactions: TIME x GROUP, $F(2, 94) = .175$, $p = .840$; TIME x PROFICIENCY, $F(2, 94) = 1.501$, $p = .228$; PROFICIENCY x GROUP, $F(1, 49) = 2.060$, $p = .158$; TIME x PROFICIENCY x GROUP, $F(2, 94) = .693$, $p = .502$.

3.4 General discussion

The goal of this chapter was twofold: first, to determine differences between IM and AH learners based on the proficiency test; second, to ascertain whether oral production developmental differences could be observed in the video retells of the IM and AH learners. This section presents first, the discussion of the results of the proficiency analysis, and second, the discussion of the results of the fluency analysis along with the findings pertaining to the effect of proficiency on different fluency measures.

3.4.1 Proficiency analysis

As was previously indicated the 27 IM learners and the 29 AH learners performed the same version of a 45-item proficiency test at the beginning and at the end of their respective learning experiences overseas (i.e., IM), and in the US (i.e., AH), and it was concluded that there were no statistically significant differences between the IM and AH groups at the beginning of each program (i.e., at Time 1), but at end only (i.e., Time 2). In addition, the analyses implied that overall both groups had improved in proficiency over the course of 6 and 14 weeks, respectively.

The fact that there were no statistical differences between the IM and AH groups at Time 1 provides strong evidence that overall, the IM and AH groups were at the same level of proficiency (i.e., grammar knowledge) before embarking on their respective learning

experiences. This is a very important aspect from a comparative standpoint given that had any proficiency differences been found at Time 1, any of the other comparisons concerning fluency and cognitive abilities could have been put into question by lack of a proficiency baseline. Except for a few exceptions such as Cubillos, Chieffo and Fan (2008), previous research on fluency has consistently overlooked the role of proficiency and has used instead tests based on oral data performance, such as OPI interviews or the ACTFL proficiency guidelines (e.g., Breiner-Sanders, Lowe, Miles, & Swender, 1999) to compare learners with different group experiences (e.g., Freed et al, 2004; Segalowitz & Freed, 2004).

Based on the proficiency scoring results for both groups at Time 1, the median score for both groups was used to determine low and high proficiency.[15] Analyses revealed that overall, high proficiency IM learners improved their grammar score more than low proficiency, however, all improvement was significant. Low proficiency IM learners experienced a statistically significant increase of proficiency between the beginning and end of the program (i.e., they scored an average 20 points at Time 1 (out of 45) and 21 at Time 2) as well as high proficiency IM learners, scored an average 28 points at Time 1 and 33 at Time 2. Low proficiency AH learners had significant improvement as they went from 19 to 21, however high proficiency AH learners did not have a statistical significant difference when they went from 27 to 28 points.

3.4.2 Fluency analysis

As was previously indicated, the IM and AH learners performed six video-retell tasks, two at each data collection time point (i.e., 2 at Time 1, 2 at Time 2, and 2 at Time 3). The use of this elicitation task contrasts with the most commonly used pre and post-test OPIs (Freed, 1995;

[15] It was seen that both IM and AH learners showed a parallel break point at 24 points (out of the maximum 45 possible points that participants could score in the proficiency test). This lead to the decision to group learners in two different proficiency groups: low proficiency learners, which would group all learners that scored between 0 and 24 points in the proficiency test at Time 1; and high proficiency learners, which in a parallel fashion, would group learners with scores between 25 and 45 points. This distribution was particularly helpful at drawing differences in the fluency and cognitive abilities analyses.

Freed, So & Lazar, 2003; Freed, Segalowitz & Dewey, 2004; O'Brien, Segalowitz, Freed & Collentine, 2007; Segalowitz & Freed, 2004; Segalowitz, Freed, Collentine, Lafford, Lazar & Díaz-Campos, 2004), informal interviews (Llanes & Muñoz, 2009), story completion task (Towell, Hawkins, & Bazergui, 1996), story telling task (Derwin, Munro, & Thomson, 2007), and controlled-per-topic sociolinguistic interviews (García-Amaya, 2008, 2009, 2010). Considering the duration of each pair of videos at each time, all learners watched 174 seconds of at Time 1, 238 seconds at Time 2, and 241 seconds at Time 3. Three main measurements were taken in each transcription: syllables produced, filled pauses produced, and seconds taken to retell the videos. Also, based on these measurements, three additional ratios were calculated, rate of speech, seconds per filled pause, and syllables per filled pause.

3.4.2.1 Amount of speech produced

The fluency analysis grouped results based on each of the measurements. First, when syllable production was analyzed, it was seen that IM learners increased their production across time up to 49% between Times 1 and 3, reaching an average of 600 syllables to retell the two videos combined at Time 3. The AH group also experienced a significant increase between Times 1 and 2, and Times 1 and 3, however, clear differences between IM and AH groups already appeared at this point. For instance, AH learners had a 29% overall increase of syllable production between Times 1 and 3, while the IM learners had a 50% increase during the same time interval. At this point, it is important to remind the reader that the highest value in syllable production at Time 3 was 299 syllables for the AH group and 600 syllables for the IM group. Although it could be argued that the increase in syllable production for the AH learners was motivated by the duration of the videos (i.e., longer videos may have involved more action and thus longer retells), it is important to note that as a group IM learners were able to double the

amount of speech of the AH learners. These differences in amount of speech produced in L2 Spanish are in line with similar findings found in other studies that also focused on the differences between the overseas IM context (i.e., 1-2 month duration including daily instruction in Spanish) and the AH context (i.e., US-based college foreign language classrooms such as Llanes & Muñoz, 2009, and García-Amaya, 2008, 2009). Other studies that have reported similar findings include research comparing US-based immersion groups with AH contexts (on L2 French, Freed, Segalowitz and Dewey, 2004 –this study also included a longer SA context), as well as others that focused only on the comparison of typically longer SA contexts (i.e., one semester) with AH contexts (Freed, 1995; Freed et al., 2003; Freed et al., 2004; García-Amaya, 2008, 2009; O'Brien et al., 2007; Segalowitz & Freed, 2004). A follow-up investigation would be necessary to see if the specific difference found between IM and AH learners for syllable production in the current investigation correlates with higher abilities of the IM group over the AH one for lexical density (i.e., amount of content and lexical words). Indeed, such a result would shed more light on the differences between the SA and AH contexts to follow up on the findings of Collentine (2004), where no statistical differences were found between the SA and AH contexts.

Another indication that IM and AH learners were following two different patterns in the development of oral skills as seen through the transcriptions of their video-retells recordings comes from measurements of amount of spoken time. The analysis showed that IM learners had a significant increase of spoken speech time between Times 1 and 3, whereas AH learners showed significant increases only between Times 1 and 2 (i.e., first 7 weeks of the experience), but not between Times 1 and 3 (i.e., the beginning and end point of the program), where more long-term gains could be claimed. Taking into account that the duration of the videos was

variable and that it kept increasing at Times 2 and 3, however, both syllable production and amount of spoken time are better understood when they are presented in the form of rate of speech in syllables per second.

This analysis revealed that IM learners had an overall significant increase of rate of speech during the first three weeks of the program (i.e., Times 1 and 2) and between the beginning and end of the six-week immersion term (i.e., Times 1 and 3). Although the rate of speech of IM learners remained faster during the last three weeks of the IM program, the overall difference between Times 2 and 3 was not statistically significant. This result indicates that the first three weeks of immersion were crucial for IM learners' rate of speech, an aspect that may be related to the constant exposure of Spanish and the limited use of English (i.e., 1-hour per week for internet access with which IM learners were granted). This important factor along with other aspects including the effects of inhibiting the L1 on the L2 are developed at length in chapter 4, where the cognitive data are also presented.

The increase in rate of speech of the IM group contrasts deeply with the overall rate of speech of the AH group, which did not increase across time, but remained stable. On average, the rate of speech of the AH learners showed a reduction of 7% from Time 1 to Time 2. Then, at Time 3, AH learners had an average rate that was almost the same as that of Time 1. In the broadest sense, this result is in line with the body of research that has shown that learners that embark on the SA experience achieve more gains in rate of speech than those that do not (i.e., Freed, Segalowitz, & Dewey, 2004; Towell, Hawkins, & Bazergui, 1996), however, in a more specific view, this result in is in line with the findings reported by Segalowitz and Freed (2004) for L2 Spanish, and Freed et al. (2004) for L2 French, where the AH groups did not show a significant improvement on rate speech between the pre-test and post-test versions of the oral

elicitation tasks. García-Amaya (2008, 2009) also explored the differences between IM and AH learners on rate of speech. Both studies concluded that IM learners were superior to AH learners for all oral measurements. The data for the two studies were based on a cross-sectional design in which the oral performance of the same groups of learners had been elicited one time through a controlled-per-topic sociolinguistic interview.

The results provided in García-Amaya (2008) were recently discussed by D'Amico (2010), where it was mentioned that a pre-test/post-test design would have been necessary to make empirically sound assertions, as it was difficult to support that the rate of speech of the learners, particularly those in the IM and SA groups, was faster due to the participation in these learning contexts. Hopefully, the highly controlled design of the results of the fluency analysis included in this dissertation will help support the claims pointed out in García-Amaya (2008, 2009) about the differences in rate of speech of our particular overseas 6-week IM learners and the 15-week US-based AH group. Moreover, the learners included in the current dissertation presented many similarities with the ones analyzed in the two previously mentioned studies (i.e., the AH group including learners with 8 years of experience with Spanish and no experience abroad, and the IM learners coming from exactly the same program and having also 6 years of experience as learners of Spanish). With the addition of these details, it is possible to surmise that the conclusions reached by García-Amaya (2008, 2009) on rate of speech were justified, at least for the IM and AH groups involved in his investigations.

Another important aspect of the current findings that warrants discussion is the effect of proficiency on fluency measures. With few exceptions in the field, most studies on second language oral development have overlooked the role of grammar knowledge on oral production (Freed et al, 2004; Segalowitz & Freed, 2004). In the current dissertation, the first proficiency

test performed by the learners in both groups served to group them into high and low levels of proficiency. This independent variable was later used in multiple statistical analyses, which indicated that there was a significant effect of PROFICIENCY on the overall rate of speech of IM learners across time, and particularly at the very beginning of the experience (i.e., Time 1) and at the very end (i.e., Time 3). In addition, the distribution according to levels of proficiency was also helpful for investigating how the fluency analysis accounted for videos 1 and 2 separately, and more specifically if the different videos had an effect on the learners' rate of speech. By means of this division, it was possible to see that low proficiency IM learners at Time 1 were the only sub-group for which the independent variable VIDEO had a minor significant effect on their rate of speech (i.e., p=.048), which disappeared at Times 2 and 3. To show why low proficiency IM learners had a slower rate of speech retelling the first video than retelling the second one is not easy, but one possibility is that some participants were confused about the objective of the task while retelling the first video, but then felt more comfortable retelling the second one. To sum up, our findings show that the IM learners significantly increased their amount of speech produced.

3.4.2.2 Hesitation phenomena

The other set of measurements included in the current dissertation pertained to the use of hesitations and more specifically to filled pauses. The IM learners in the current dissertation showed a significant increase of filled pause production across time, particularly during the first 3 weeks (i.e., between Times 1 and 2) and across time (i.e., between the beginning and end of the program). Although it could be assumed that the increase of filled pauses was due to the production of longer turns across time, the analysis of the rate of seconds per filled pause, which indicated the number of seconds that pass between filled pauses, revealed that IM learners had a

significant decrease of rate of seconds per filled pause, or in other words, that IM learners reduced the amount of time between filled pauses across time (i.e., IM learners significantly decreased their rate of filled pauses per second across time between Times 1 and 2, and between Times 1 and 3). The analysis for rate of syllables per filled pause revealed that this rate also diminished across time (i.e., less syllables between filled pauses) in a significant fashion (particularly between Times 1 and 2 and between Times 1 and 3).

The reasons for this increase of filled pause production were analyzed extensively. First, it was verified that there was no effect of VIDEO on each of the two rates (i.e., rate of seconds per filled pause and rate of syllables per filled pause) at any given time. The role of PROFICIENCY was also investigated, and surprisingly, when the differences for the two proficiency subgroups were analyzed, it was seen that each subgroup followed a very different trend. While low proficiency IM learners decreased their rate of seconds per filled pause, high proficiency IM learners increased it. Although the statistical analysis did not show significance for the effect of proficiency at either time point (as it was point out before in the results section, this was probably because there were very few tokens as well as some learners that did not produce any filled pauses), the tendencies for each group were straightforward and indicated that high proficiency IM learners produced less filled pauses across time and that low proficiency IM learners produced more. In the same fashion, high proficiency IM learners also increased their rate of syllables per filled pause across time, while IM low proficiency learners kept decreasing it. Obviously, the results for low and high proficiency IM learners for rate of seconds per filled pause and rate of syllables per filled pause mirror each other because the amount of spoken time in seconds and the amount of oral production in syllables are well correlated. However, this set of findings adds to previous research based on overseas IM contexts similar to the one studied in

the current dissertation (García-Amaya, 2010; García-Amaya & Filgueras Gómez, 2010; Czerwonka & García-Amaya, 2011). Although previous studies did not examine the role of proficiency on filled pause production, they did show an overall reduction of the production of filled pauses as well as on the rate of seconds per filled pause in their different oral elicitation tasks. The findings of the current dissertation for the IM group do not necessarily contradict previous findings on the use of filled pauses; on the contrary, they offer new insights into the role of hesitation in video-retells, another type of task that had not been used in previous second language oral fluency research.

For AH learners, it was found that these participants also showed a similar average filled pause production across time independently of the video and/or its duration. In fact, the rate of seconds per filled pause also showed a very stable production of filled pauses across time with no significant effect of TIME. Although proficiency was not a significant factor either, overall, high proficiency AH learners had lower rates of seconds per filled pause than their low proficiency counterparts at Time 1 (i.e., 30 and 13 seconds per filled pause respectively), continued that trend at Time 2 (i.e., 25 and 12 seconds per filled pause respectively, and finally balanced their production at Time 3 (27 and 26 seconds per filled pause respectively). Again, it is important to bear in mind that there was not a significant effect of PROFICIENCY on the RATE OF SECONDS PER FILLED PAUSE at any data collection time point.

AH learners did not show an effect of TIME or PROFICIENCY on the rate of syllables per filled pauses either. However, the overall raw numbers did show and increasing tendency towards similar rates of syllables per filled pause. Only the difference at Time 2 between low and high AH proficiency learners approached significance (i.e., p=.089), however, at Time 3 the

overall rates for low and high proficiency AH learners were very similar (49 and 55 syllables per filled pause respectively).

Traditionally, the reduction across time of filled pauses has been interpreted as a sign of fluency improvement (Freed, 1995; García-Amaya, 2009; 2010, Lennon, 1990; Llanes & Muñoz, 2009; Riggenbach, 1991; Towel et al., 1996). Although each definition of fluency in itself may view the production of filled pauses in different way, the consensus is that filled pauses break the flow of 'smoothness in speech'. The data from video-retells in the current dissertation show that although recent research has shown that participants with experience in the abroad context may produce less filled pauses when they are presented with the same elicitation task at the beginning and end of their contextualized learning experiences, that may not be the case when the post-test task, although fairly similar, is not exactly the same. In addition, an increase of filled pauses across time does not necessarily mean lack of improvement in oral skills. On the contrary, it may indicate that learners are using filled pauses to gain time while they process more complex syntactic structures, beyond the independent clause, that require for instance the use of subjunctive verbal forms (i.e., nominal, adverbial, and adjectival clauses). In this regard, García-Amaya (2010) carried out an analysis of the relationship of filled pauses to clause types and concluded that use of filled pauses is not random but follows very specific trends related to the position before or within a clause, a tendency that is also statistically influenced by the clause type (i.e., independent clauses, sub-clause independent units, nominal clauses, adjectival clauses, and/or adverbial clauses).

As was noted on chapter 1, Clark and Fox-Tree (2001) expressed the importance of considering filled pauses such as *uhs* and *uhms* as real words because speakers produce them at the beginning of clauses much more frequently than within clauses. Also, Segalowitz (2010) has

expressed that learners' use of filled pauses and pauses is a necessary component of the development of oral speech production in a second language. In the same fashion, it is important to say that future research should analyze the L1 productions of learners. As has already been noted by Fox Tree (2001, 2002) and Clark and Fox Tree (2002), the use of filled pauses such as *ums* and *uhs* may have a communicative role in L1 speech (taking the floor, delays, errors, assertions, reformulations, etc.), and it would be very relevant for researchers in L2 speech production to investigate if these aspects transfer to the L2.[16]

3.5 Conclusion

In this chapter, the results for the fluency analysis of the video-retells were presented. Clear differences were seen between the IM and AH learners in all measurements. From a strictly statistical point of view, IM learners increased their spoken time, their production of syllables and their rate of speech across time, whereas AH learners remained stable across the three time points. Also, IM learners showed a significant increase in their production of filled pauses seen through the rate of seconds per filled pause and the rate of syllables per filled pause. On the contrary, AH learners also kept these two rates at similar levels across time.

As was previously mentioned all videos used in the stimuli were different and thus the significant improvement of the IM group cannot be associated with the fact that they were asked to retell the same video or that they had memorized their plots. In addition, the measurements of rate of speech reported in this chapter were calculated from two different videos, which were associated in terms of characters, but always developed a different plot. Thus, the increase

[16] Previous research with native speakers of Spanish (García-Amaya, 2009) has shown that native speakers of Spanish barely produce filled pauses, and as Fox tree and Clark have shown, it is possible that filled pauses are more frequent in L1 English than in L1 Spanish. Future research should consider a thorough analysis of native speakers of both languages learning English and Spanish as second languages in order to draw sound conclusions about the use of these structures.

observed for the IM learners in the six weeks between Time 1 and Time 3 truly represents a developmental improvement in their oral skills.

Analysis of the oral production data in terms of PROFICIENCY also indicated that low and high proficiency learners showed no statistical significant differences in their rate of seconds per filled pauses and rate of syllables per filled pause. As for rate of syllables per filled pause it was also seen that whereas high proficiency IM learners continued increasing their rates (i.e., more syllables between filled pauses), low proficiency IM learners continued decreasing their rates, thus producing more filled pauses. Future analysis of these data will include other fluency measures such as pausing time, articulation rate, reformulations, repetitions and rate speech of adjusted (that does not include neither reformulations nor repetitions). These data would also help to understand how these other measures correlate with the cognitive data presented in chapter 4.

The decision to plan a design that allowed collecting oral data at three points in time also allowed for unique insight into the development of oral production for these two groups of learners. Collecting these types of data at three time points provided a more detailed analysis of the development of second language oral production. Specific aspects of interlanguage development such as the importance of the first part (i.e., first three week) of the IM context for the advancement of rate of speech have been overlooked in previous research, where data were collected with a pre and post-test oral elicitation task (Freed, 1995; Freed et al, 2003; Freed et al., 2004; Isabelli-García, 2003; Juan-Garau and Pérez-Vidal, 2007; Lennon, 1990; Llanes & Muñoz, 2009; O'Brien et al., 2007; Segalowitz et al., 2004; Towell et al., 1996; Towell, 2002). Another advance in the design of this dissertation came from collecting data in situ, instead of before and after the abroad experience (Segalowitz & Freed, 2004). Although there is no doubt that the oral

gains after participating in a study abroad experience remain once the experience is over, it cannot be disputed that the learner's language mode changes as soon as or very quickly after he or she is in contact with the native language on a daily basis. It is believed that the studies that did not clarify this aspect as well as the length of time between the return from abroad and the moment when the data were collected have a methodological shortcoming, as important developmental aspects of oral production were never reported and different stages that took place in the specific language-learning contexts were missed.

CHAPTER 4: COGNITIVE ABILITIES FOR IM LEARNERS

4.1 Introduction

During the last 30 years, the questions that researchers in the field of L2 fluency have addressed pertain to the following topics: fluency improvement, operationalized as the automaticity (or facility) of delivery of speech that learners reach after participating in an immersion experience abroad (Lennon, 1990, Freed, 1998; Freed, Segalowitz and Dewey, 2004; Segalowitz and Freed, 2004); the production of disfluency markers such as hesitation phenomena (i.e., silent and filled pauses), self-repairs, and repetitions (Averart and Nation, 1991, Fillmore, 1979; Kormos, 2000; Nation, 1989; Raupach 1980; Riggenbach, 1991; see review in Koponen & Riggenbach, 2000); and the influence of these phenomena on the perception of accentedness, comprehensibility or intelligibility of L2 speech by both naïve and professional raters (Derwing & Munro, 1997; Flege, Munro, & MacKay, 1995, Munro & Derwing, 1995, 2001). Although these studies have provided relevant results that modulate our understanding of interlanguage oral development, it is still necessary to investigate the relationship between the characteristics of learners' oral production and the cognitive processes that underlie speech production.

To this end, this chapter analyzes the development of cognitive abilities of the 27 IM learners analyzed in previous chapters of this dissertation. The reason to include this group only is that the IM learners were the only group that showed improvement in oral fluency during their learning context and thus, the only one that rendered an interest to understand if other cognitive abilities had also changed or developed during the 6-week immersion program. As discussed in chapter 1, the three cognitive abilities under investigation are lexical access, lexical retrieval, and attention control. These have been proposed to be closely related to lexical access and fluency in earlier studies (DeKeyser, 1991; Silverberg and Samuel, 2004). In order to measure each of these

cognitive abilities considered in this dissertation, IM learners performed different cognitive tasks that are widely used in the psycholinguistics literature to access each of the abilities under investigation. Two semantic classification tasks (i.e., SCT), one in English and one in Spanish, were devised to measure lexical access (Segalowitz and de Almeida, 2002; Ding, Perry, Peng, Ma, Li and Xu, 2003, Thompson-Schill and Gabrielli, 1999; Young, 1992). A Spanish version of a picture-naming task (i.e., PNT) was designed to measure lexical retrieval (Kroll, 1993; 1994; Kroll and Tokowicz, 2001; Kroll and Sunderman, 2003; Wang, Xue, Chen, Xue and Dong, 2007). Finally, an English version of an auditory discriminatory task was implemented to measure attention control, as in Darcy, Park and Yang (2011). While the same versions of the English SCT and the English attention control tasks (i.e., ACT) were performed at the beginning and at the end of the IM program (i.e., pretest and posttest versions of these two tasks), three versions of the Spanish SCT and of the Spanish PNT were administered at the beginning, middle and end of the IM program. The analyses focused on overall performance in each cognitive task, particularly to the effect of Time and Proficiency on RT, as well as individual differences across participants. The latter aspect is highly relevant to this particular investigation as different patterns across different cognitive tasks are expected among IM participants.

Chapter four is organized as follows: first, sections 4.2, 4.3, 4.4, and 4.5 present and discuss accuracy and response time results of the Spanish SCT, English SCT, Spanish PNT, and English ACT, respectively. Section 4.6 includes the general discussion and conclusion.

4.2 Results of the Spanish SCT for accuracy and speed (Spanish lexical access task)

4.2.1 Accuracy results

This section introduces the accuracy results for the IM group for the Spanish SCT (for a review of the specifics of this task, see section 2.3.3.2 in Chapter 2). Only the results for the

cognitive abilities performed by the IM learners were included in this dissertation. The accuracy results for the Spanish classification task indicate that across time, participants matched correctly a total of 71.2% of the stimuli as either living or non-living, while 28.8% of their responses were incorrectly matched. Table 4.1 shows accuracy rates, including correct responses per data collection time point (i.e., Time 1, Time 2 and Time 3), as well as accuracy rates for NEW and OLD stimuli. Overall, participants showed an increase of accuracy from Time 1 to Time 2 and then remained stable at Time 3. When the accuracy rates were broken down between NEW and OLD stimuli, it was seen that accuracy rates for NEW stimuli increased drastically between Times 1 and 2 and slightly increased at Time 3, while for OLD stimuli, accuracy was slightly higher at Time 2 than at Time 3 but always above the 80% threshold. Moreover, the difference in accuracy rates between NEW and OLD stimuli was minor (i.e., 81.4% to 82.1%, respectively) at Time 3.

In order to investigate the effect of TIME on ACCURACY, a series of Generalized Linear Models, a statistical procedure that allows the analysis of binary independent variables, were conducted. The first model analyzed the effect of TIME on ACCURACY for all stimuli (i.e., both NEW and OLD stimuli) and revealed a significant effect between Times 1 and 2, Wald Chi-Square = 1664.575, p=.000; Time 1 and 3, Wald Chi-Square = 1664.575, p=.000; but not between Times 2 and 3, Wald Chi-Square = 1664.575, p=.079. This indicates that when all stimuli are considered, the largest change in accuracy appeared between the first three weeks of the immersion stay. Next, a second model analyzed the effect of TIME on ACCURACY for NEW stimuli only and it showed that there was a significant effect of TIME between Times 1 and 2, Wald Chi-Square =846.359, p=.000, Times 1 and 3, Wald Chi-Square = 856,359, p=.000, and also Times 1 and 3, Wald Chi-Square =846.359, p=.000, which shows that IM learners kept

improving their accuracy at each TIME for items that had not been shown to them in previous versions of the Spanish SCT (i.e., at either Time 2 or Time 3). The third model was run to analyze the effect of TIME on ACCURACY for OLD stimuli only, and it showed that IM learners were statistically more accurate with OLD stimuli at Time 2 than at Time 3, Wald Chi-Square = 18.074, p=.000.

Table 4.1. Spanish SCT accuracy and SD for IM learners for correct responses only

TASK	TIME 1	TIME 2		TIME 3	
ACC (%)	52.2%	84.1%		81.8%	
STIMULI	**NEW**	**NEW**	**OLD**	**NEW**	**OLD**
ACC (%)	52.2%	77.6%	89.3%	81.4%	82.1%

The following section will focus on reaction times (RTs) in milliseconds, calculated based on correct answers only of the Spanish SCT. Group averages are presented first according to proficiency; I then report individual results.

4.2.2 Global speed results

This section introduces the speed results (i.e., RTs in milliseconds) of the IM learner group for Spanish SCT. The purpose of this task was to investigate if the speed of lexical access with which IM learners accessed Spanish words in their mental lexicon varied during the course of the six-week IM program. Figure 4.1 shows that the mean RT for the Spanish SCT was faster at Times 2 and 3 than at Time 1, and a Linear Mixed Model (all statistical analysis included in this chapter were calculated with this statistical procedure) revealed a significant effect of Time on RT between Times 1 and 2, $F(2,3176) = 29.38$, $p=.000$, between Times 1 and 3, $F(2,3176) = 29.38$, $p=.000$, but not between Times 2 and 3, $F(2,3176) = 29.38$, $p=.336$.

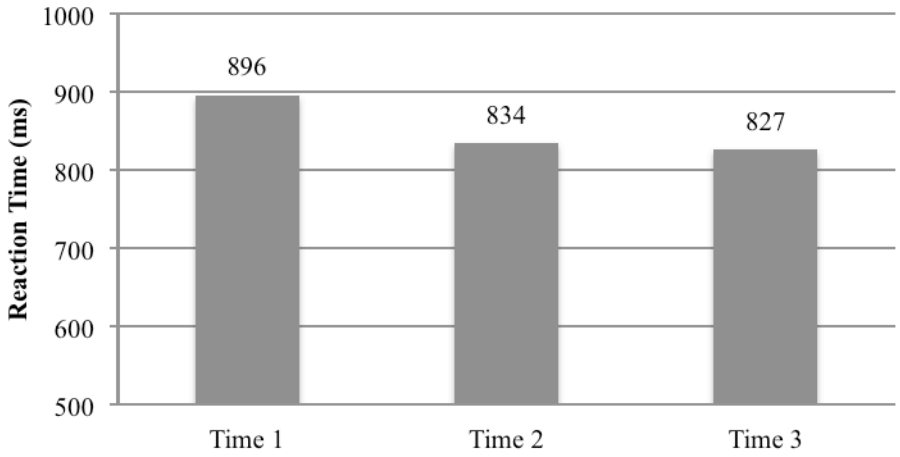

Figure 4.1. RT for new words only for IM learners as a function of Time

The significantly faster RTs along with higher accuracy rates on new words reveals an underlying cognitive development, by which L2 learners of Spanish, overall, access L2 words in their mental lexicon faster in six weeks.

Next, the effect of PROFICIENCY on RT was also analyzed, and it was found that there was not a overall significant effect of PROFICIENCY on RT, $F(1,25)=.014$, $p=.908$. In fact, when the role of PROFICIENCY on RT was analyzed at each time point individually, it was found that there was not a significant effect at any time point, either Time 1, $F(1,25) = .351$, $p=.559$; Time 2, $F(1,25) = .084$, $p=.774$; or Time 3, $F(1,25) =.060$, $p=.808$. However, there was a significant effect for the interaction of TIME and PROFICIENCY on RT, $F(2,3176) = 3.546$, $p=.029$, which implies that low and high proficiency learners evolve differently over time. Figure 4.2 presents the results of the Spanish SCT separated according to the results obtained in this task by high and low proficiency IM learners.

The data indicate that the average RT of low proficiency IM learners became 10% faster between Times 1 and 2 and remained at a similar speed at Time 3. There was a significant effect of TIME between the beginning and end of the program (i.e., Times 1 and 3), $F(2,1618) = 22.92$, $p=.000$, and also during the first three weeks (i.e., between Times 1 and 2), $F(2,1618) = 22.92$,

p=.000. However, there was no significance during the last three weeks of the program (i.e., between Times 2 and 3), F(2,1618) = 22.92, p=1.000. The results obtained by high proficiency IM learners (i.e., those learners that scored between the range of 25 and 45 points in the proficiency test administered prior to the beginning of the program) show that these learners became significantly faster across time, as shown by a significant effect of TIME between the beginning and end of the program (i.e., between Times 1 and 3), F(2,1547) = 11.26, p=.000. There was a statistically significant decrease in speed for high proficiency IM learners during the last three weeks of the program (i.e., between Times 2 and 3), F(2,1547) = 11.26, p=.018. However, this significant effect did not appear during the first three weeks of the program (i.e., between Times 1 and 2), F(2,1547) = 11.26, p=.498.

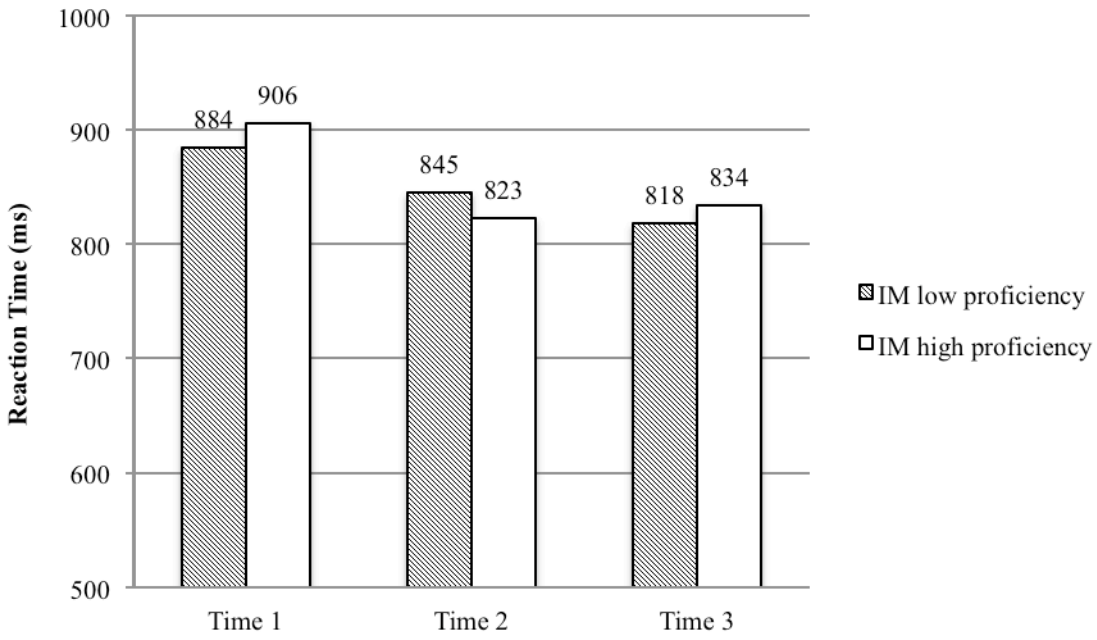

Figure 4.2. Spanish SCT average RT for low and high proficiency IM learners across time

To summarize, while high proficiency learners had significantly slower RTs between Times 2 and 3, low proficiency learners had significantly faster RTs between Times 1 and 2. This would indicate that for low proficiency IM learners the increase of speed accessing Spanish

words in their mental lexicons was quick during the first three weeks, while for the high proficiency learners, a decrease appeared during the last three weeks. This is very important as it proves the significant PROFICIENCY and TIME interaction previously reported for the IM learner group.

4.2.3 Speed results for new and old stimuli

In this section, the overall RTs obtained by all IM learners are investigated in detail by comparing learners' performance on NEW and OLD stimuli. The design of the task included different degrees of new and old stimuli at Times 2 and 3 (i.e., at Time 2, 50% of new stimuli and 50% of old stimuli from Time 1; and at Time 3, 50% of new stimuli, 25% of old stimuli from Time 1 and 25% of old stimuli from Time 2) to test the hypothesis that participants would be able to access old items faster than new ones because they had been familiar with them. Recall that there were only three weeks between data collection time points.

Figure 4.3 shows the RTs for NEW and OLD stimuli across time for the Spanish SCT. As for OLD stimuli at Time 2 and 3 only, the statistical analysis indicated that there was no significant effect of TIME on RT, $F(1,2082) = .007$, $p=.933$. Then, when OLD and NEW stimuli were compared, a significant effect of NEW vs. OLD on RT was observed at Time 2 indicating that, as was expected, NEW stimuli were responded to slower than OLD stimuli, $F(1,1887) = 9.503$, $p=.002$. On the other hand, a similar effect was not found at Time 3, $F(1,1881) = .000$, $p=1.000$. Thus, the data from the Spanish SCT indicate that overall, by the end of the immersion program, all IM learners accessed new and old stimuli at virtually the same speed.

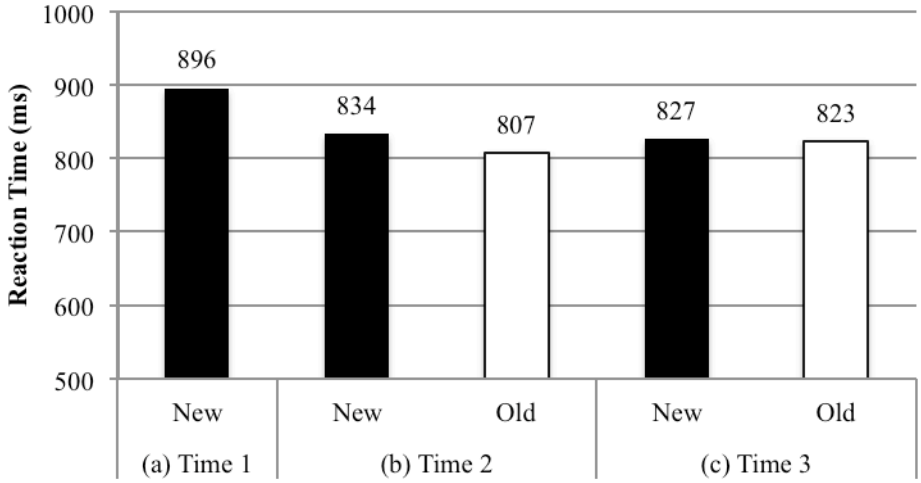

Figure 4.3. Spanish SCT average RT for NEW and OLD stimuli for the IM learner group as a function of TIME

Figure 4.4 shows the RT results for NEW and OLD stimuli separated according to levels of proficiency. The statistical analysis indicated that the there was no significant effect of TIME on OLD stimuli for either low proficiency IM learners, $F(1,1066) = 1.714$, $p=.191$, or high proficiency IM learners, $F(1,1007) = 1.891$, $p=.169$, which indicates that overall, neither subgroup accessed OLD words faster at Time 2 than at Time 3. Next, the effect of NEW vs. OLD on RT was analyzed for each subgroup, and it was found that there was a significant effect at Time 2 for both low proficiency IM learners, $F(1,939) = 5.22$, $p=.022$, and also high proficiency IM learners, $F(1,939) = 4.50$, $p=.034$. This result indicated that both subgroups were faster with OLD stimuli than with NEW stimuli at Time 2. Moreover, a significant effect of NEW vs. OLD on RT was also found at Time 3 for low proficiency IM learners, $F(1,979) = 7.27$, $p=.007$, which showed that low proficiency IM learners were also faster with OLD stimuli at Time 3. However, for high proficiency IM learners, a significant effect was found but in the opposite direction, $F(1,892) = 7.93$, $p=.005$, indicating that overall high proficiency IM learners were statistically faster with NEW stimuli at Time 3 (815ms vs. 851ms, respectively).

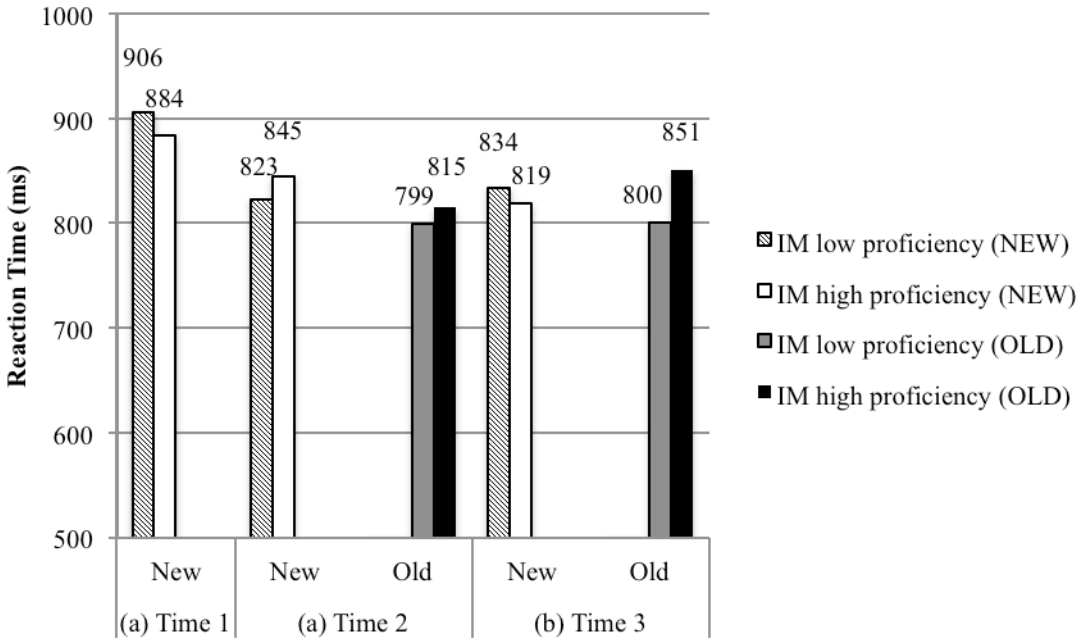

Figure 4.4. Spanish SCT average RT for NEW and OLD stimuli for IM learners separated according to levels of proficiency

Once more, when the data for the Spanish SCT were separated according to levels of proficiency, differences could be observed. First, at Time 2, low and high proficiency learners were statistically faster with OLD stimuli compared to NEW stimuli. At Time 3, low proficiency learners only were also significantly faster with OLD stimuli. On the contrary, high proficiency learners resulted statistically faster with NEW stimuli compared to OLD stimuli. This result was highly unexpected since for high proficiency learners a larger mean difference between OLD stimuli at Times 2 and 3 than the one found between NEW and OLD stimuli at Time 3 was not significant. It is possible that the latter is a spurious significance and should interpreted with caution.

4.2.4 Individual speed results for new items

Individual differences are pervasive in all domains of second language acquisition and are also highly revealing of the developmental processes of interlanguage development (Dörnyei, 2008; Skehan, 1999; Segalowitz, 1997; Cohen and Dörnyei, 2002; Dörnyei and Skehan, 2003).

As argued by Larsen-Freeman (1991), an individual differences analysis goes beyond group categorizations and focuses on the differential aspects of L2 learners and their individual developmental sequences.

Figure 4.5 shows the RT calculated for correct responses only for low proficiency IM learners across the three data time points on this task. The figure includes a legend for each of the five possibilities of speed development that low proficiency IM learners adopted in this task during the six weeks of the program. Each type is represented with a combination of three letters (i.e., F-F-F, F-S-F, F-S-S, S-F-F, and S-F-S). The first letter of each combination indicates if the learner was faster (i.e., F) or slower (i.e., S) at Time 2 compared to Time 1; the second letter indicates faster or slower for Time 3 with respect to Time 2; and finally, the third letter marks if the overall speed at Time 3 was faster or slower than at Time 1. For example, F-S-S learners are faster at Time 2 than at Time 1, slower at Time 3 than at Time 2, and overall slower at Time 3 than at Time 1. In the sample of 14 low proficiency learners there were 2 F-F-F learners, 7 F-S-F learners, 2 F-S-S learner, 1 S-F-F learner, and 2 S-F-S learners.

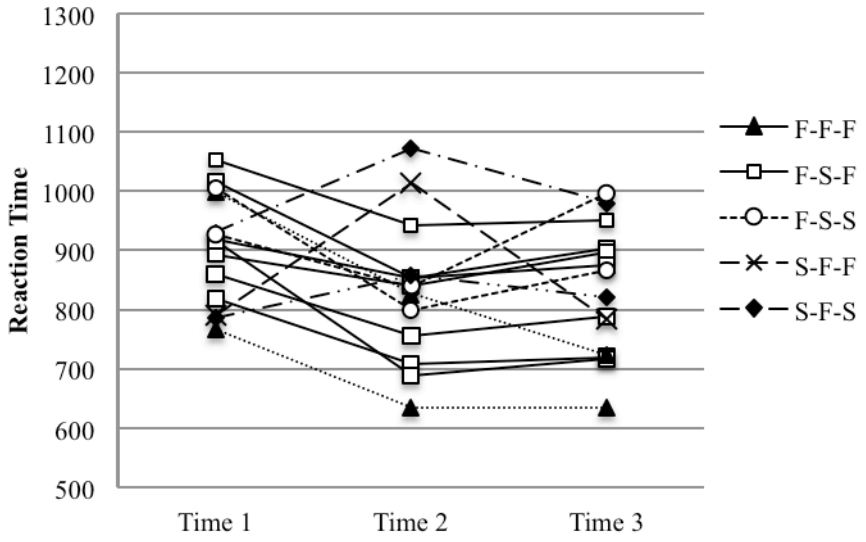

Figure 4.5. Spanish SCT individual differences for IM low proficiency learners as a function of time

Figure 4.6 presents the analysis of individual differences for high proficiency IM learners. As the legend indicates, the patterns followed for high proficiency learners are the same that were shown in Figure 4.7 (i.e., F-F-F, F-S-F, and S-F-F) with the exception of S-S-S, a trend followed by high proficiency IM learners only and which indicates that participants were slower at Time 2 than at Time 1, at Time 3 than at Time 2, and also slower at Time 3 than at Time 1. Out of the 13 high proficiency learners included in the sample, there were 8 F-F-F learners, 1 F-S-F learner, 2 S-F-F learners and 2 S-S-S learners.

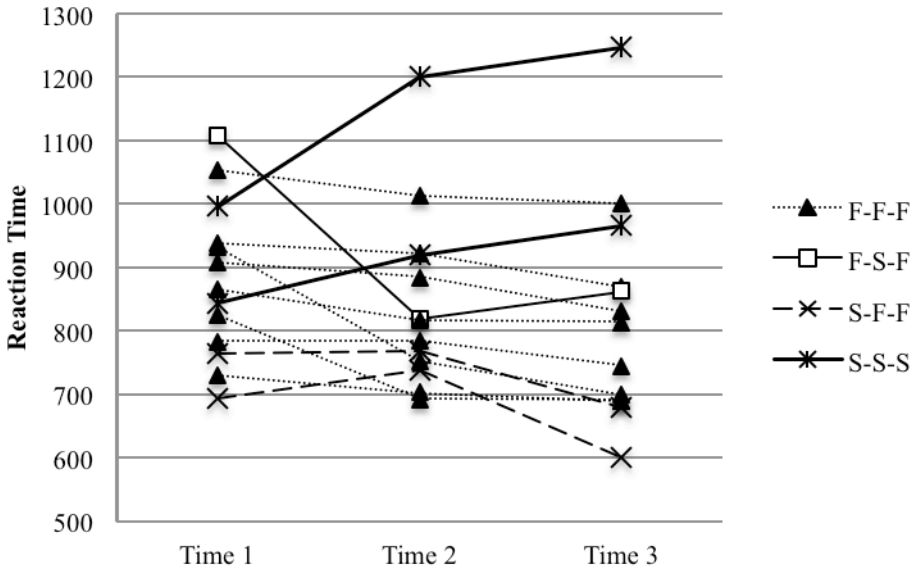

Figure 4.6. Individual differences for IM high proficiency learners for the SCT based on Time.

Although these trends may give the impression of high variability, not all results were statistically significant, and thus, these ranges were very different from the figures reported in other studies that include similar task types (cf. Segalowitz and Freed, 2004). Table 4.2 shows the individual RTs for each IM learner at each Time, as well as the level of significance between Times 1 and 2, Times 2 and 3, and Times 1 and 3. The table indicates that eight participants show a statistically significant difference between Times 1 and 2. There are only three participants that show a statistically significant difference between Times 1 and 2, two of which

are also the only two participants that show a significant difference between Times 2 and 3. As for the differences between Times 1 and 3, nine participants are significantly faster between Times 1 and 2, while three participants are significantly slower between Times 1 and 3.

In sum, out of 27 participants, eight are significantly faster between Times 1 and 2, whereas two are significantly slower. Between Times 2 and 3 three learners only are significantly faster. Between Times 1 and 3, a total of ten participants were statistically faster while only two were significantly slower. When we factor in the role of proficiency, between Times 1 and 3, only none of the low proficiency learners was significantly slower, while five of them were significantly faster. On the contrary, while five of the high proficiency learners were significantly faster, two of them were significantly slower.

Table 4.2. Individual differences of IM proficiency learners for the Spanish SCT and significance.

PROF	PART.	T1	T2	T3	T1-T2	T2-T3	T1-T3	SIGNIFICANCE T1-T2	SIGNIFICANCE T2-T3	SIGNIFICANCE T1-T3
LOW	1	916	688	716	F***		F***	F(1,81) = 16.29, p=.000	F(1,53) = .516, p=.476	F(1,80) = 12.58, p=.001
LOW	2	1016	853	904	F***		F*	F(1,82) = 12.22, p = .001	F(1,59) = .880, p=.352	F(1,84) = 4.89, p=.030
LOW	4	927	839	996				F(1,51) = 2.144, p=.149	F(1,44) = 3.45, p=.70	F(1,55) = 1.48, p= .230
LOW	8	894	841	897				F(1,73) = .957, p=.331	F(1,58) = 1.29, p=.260	F(1,82) = .007, p=.935
LOW	9	919	854	875				F(1,73) = .154, p=.696	F(1,47) = .079, p=.780	F(1,80) = .59, p=.443
LOW	11	930	1072	980				F(1,66) = 3.52, p=.065	F(1,53) = .921, p=.341	F(1,71) = 1.12, p=.294
LOW	12	791	1014	784	S*	F***		F(1,56) = 5.49, p=.023	F(1,36) = 11.97, p=.001	F(1,79) = .29, p=.586
LOW	13	766	634	634	F*		F***	F(1,75) = 6.68, p=.012	F(1,55) = .243, p=.624	F(1,80) = 10.91, p=.001
LOW	14	787	858	832				F(1,75) = 1.03, p=312	F(1,52) = .243, p=.624	F(1,78) = 1.75, p=.190
LOW	17	1054	943	952				F(1,72) = 3.83, p=.054	F(1,53) = .096, p=.757	F(1,88) = 3.41, p=.068
LOW	21	998	826	723	F**	F**	F***	F(1,76) = 7.23, p=.009	F(1,49) = 8.57, p=.005	F(1,81) = 33.64, p=.000
LOW	23	819	708	720	F*			F(1,71) = 4.30, p=.042	F(1,52) = .017, p=.898	F(1,75) = 2.93, p=.091
LOW	25	860	755	788				F(1,78) = 3.91, p=.052	F(1,58) = .026, p=.872	F(1,85) = 1.09, p=.300
LOW	27	1004	799	866	F***		F**	F(1,69) = 10.66, p=.000	F(1,52) = 2.85, p=.097	F(1,78) = 6.30, p=.014
HIGH	3	907	884	830				F(1,72) = .129, p=.721	F(1,55) = 1.52, p=223	F(1,82) = 3.23, p = .076
HIGH	5	843	919	966			S**	F(1,74) = 1.70, p=.196	F(1,48) = .161, p=.690	F(1,76) = 5.92, p=.017
HIGH	6	1053	1013	1000				F(1,84) = .022, p=.883	F(1,63) = .82, p=.367	F(1,90) = .39, p=.536
HIGH	7	730	703	690				F(1,80) = 1.20, p=.277	F(1,55) = .59, p=.446	F(1,83) = 1.36, p=.247
HIGH	10	693	738	601		F***	F**	F(1,54) = 3.49, p=.067	F(1,51) = 40.17, p=.000	F(1,62) = 9.007, p=.004
HIGH	15	996	1200	1248	S**		S***	F(1,62) = 9.58, p=.003	F(1,45) = .152, p=.699	F(1,69) = 17.69, p=.000
HIGH	16	764	769	679			F*	F(1,74) = .02, p=.893	F(1,55) = 2.00, p=.162	F(1,80) = 4.60, p=.035
HIGH	18	865	816	814				F(1,80) = .89, p=.348	F(1,62) = .010, p=.919	F(1,85) = .96, p=.331
HIGH	19	939	922	870			F*	F(1,75) = 1.13, p=.290	F(1,55) = 2.85, p=.097	F(1,81) = 4.68, p=.033
HIGH	20	783	784	745				F(1,81) = .12, p=.732	F(1,54) = .82, p=.368	F(1,83) = .51, p=.475
HIGH	22	1108	819	863	F***		F***	F(1,71) = 11.74, p=.001	F(1,50) = .38, p=.540	F(1,80) = 10.93, p=.001
HIGH	24	834	692	690	F*		F**	F(1,77) = 5.29, p=.024	F(1,50) = .005, p=.942	F(1,83) = 10.14, p=.002
HIGH	26	932	752	700	F**		F***	F(1,73) = 5.70,p=.020	F(1,54) = 2.98, p=.090	F(1,85) = 26.05, p=.000

* = p ≤ .05, ** = p ≤ .01, *** = p ≤ .001

4.2.5 Summary and discussion

The Spanish version of the SCT used in this dissertation was initially modeled after the SCT used in Segalowitz and Freed (2004), however a number of methodological and experimental changes were adopted in this task. First, instead of collecting the data before and after participants had engaged in their SA experience, the data presented here were collected *in situ* once participants had arrived to Spain. Second, while Segalowitz and Freed used the same task as pre and post-test, the present SCT had three versions that were presented to participants every three weeks (i.e., a total of three times).

In the same way as the SCT of Segalowitz and Freed (2004), the version reported here also allowed for accuracy and speed results for the IM learners. Collecting data at three points in time proved an efficient design as it allowed to see that in three weeks, IM learners went from 52% accuracy to surpassing 80%, a percentage that they maintained at Time 3. When NEW and OLD stimuli were brought into picture, the accuracy analysis indicated that at Time 2, IM learners were more accurate with OLD stimuli (89.3%) than with NEW ones (77.6%), a significant difference between old and new stimuli that disappeared at Time 3 with 82.1% accuracy for OLD stimuli and 81% accuracy for NEW stimuli.

Focusing now on the speed measurements of the Spanish SCT, the RTs revealed that independently of being exposed to new stimuli, IM learners became significantly faster accessing Spanish words in their mental lexicon during the first three weeks of the program (i.e., Times 1 and 2), and also across the whole program duration (i.e., Times 1 and 3). When the data were separated according to levels of proficiency, high proficiency learners were overall faster than low proficiency learners, and both high and low proficiency learners were significantly faster between Times 1 and 2 and Times 1 and 3, which would indicate that in this particular learning

context, at least during the first three weeks, the multiple benefits of the IM context (i.e., extensive availability of input, instruction, exposure to a variety of native speakers, and inhibition of the L1 among other factors) facilitated IM learners' speed at accessing Spanish words in their mental lexicon. Further analysis examined the RT times for NEW and OLD stimuli at Times 2 and 3. Overall, IM learners were slower with new stimuli than with old stimuli at Time 2 and at Time 3, and these results were statistically significant.

When these data were separated according to high and low proficiency learners, two important findings were revealed. First, low proficiency learners followed a very similar trend as the one previously reported, with significantly slower RTs for new stimuli than for old stimuli. Second, although high proficiency IM learners also followed this trend at Time 2, they became faster with new stimuli than with old stimuli at Time 3. Indeed, they produced their overall fastest RT in this task with this particular set of words that had never been shown to them before. The significant result indicates that after six weeks of immersion, high proficiency IM learners as a group were able to access new stimuli faster than Times 1 and 2, and according to the statistical analysis it could be interpreted that overall they were also able to inhibit or attenuate old stimuli that had been presented to them in previous versions of the task. Interestingly, significant differences for the RT of old stimuli at Times 2 and 3 did not appear for either of the IM learner subgroups in general. The overall interpretation is that the significance between NEW and OLD stimuli for high proficiency learners may be rather spurious and therefore not fundamental for this analysis.

The next step in the analysis dealt with the individual data, which was first addressed according to levels of proficiency and afterwards per participant. The proficiency subdivision indicated that while five of the 13 high proficiency learners were significantly faster between the

beginning and end of the program (i.e., F-F-F) and only 2 were significantly slower. On the other hand, out of 14 low proficiency learners five were significantly faster and none of them were significantly slower. However, the most important piece of information regarding the individual data is that the significance occurs mostly between the first three weeks in the program where six of the low proficiency learners were significantly faster and one was significantly slower at Time 2 than at Time 1, whereas only three high proficiency learners were significantly faster and only one was significantly slower at Time 2 than at Time 1. These results would support the idea that the low proficiency learners received most benefit early, an aspect that also holds true for fluency measures such as rate of speech. As was shown in Chapter 3, there was a significant effect of PROFICIENCY on RATE OF SPEECH at Time 1, $F(1,25) = 6.105$, $p=.021$. However, this difference was not observed at Time 2, $F(1,25) = 2.883$, $p=.102$, and it was barely significant at Time 3, $F(1,25) = 4.448$, $p=.045$. Indeed, low proficiency IM learners achieved a significantly faster rate of speech between Times 1 and 2, $F(2,26) =31.122$, $p<.001$, Times 2 and 3, $F(2,26) = 31.122$, $p=.035$, and also between Times 1 and 3, $F(2,26) = 31.122$, $p<.001$. However, high proficiency learners only showed a significant effect between Times 1 and 3, $F(2,24), = 11.562$, $p<.001$. Moreover, when looking again at the results of the Spanish SCT between Times 2 and 3, it was found that only two low proficiency learners and only one high proficiency learner were significantly faster at Time 3 than at Time 2. If we add this piece of information to what has been previously mentioned about the Spanish SCT as well as rate of speech, it would seem that particularly for low proficiency learners the main changes at the cognitive level actually happened during the first three weeks and not within the second part of the program.

Additionally, it is important to mention that there were no differences in accuracy between new and old stimuli at Time 3 (81.4% accuracy for NEW stimuli and 82.1% accuracy

for OLD stimuli). The individual analysis carried out for the Spanish SCT indicates that in order to analyze the development of lexical access thoroughly, it is important to look at overall RTs for the whole group and also to pay attention to the difference between the two levels of proficiency. The individual analysis provides a new dimension to show that although overall, the IM group seemed to become faster with the Spanish SCT, only nine of them did so with a statistically significant difference. In addition, three IM learners followed an opposite trend and had statistically significant slower times at the end of their experience than at the beginning.

Although the results of the Spanish SCT in this dissertation do not contradict those of Segalowitz and Freed, overall, both the SA learners in their study over four months, and the IM learners investigated in this dissertation over six weeks, were able to achieve faster RTs at the end of their respective learning contexts. What must be left for future research is to determine whether it is the abroad experience itself (which in this study lasted six weeks and had very particular components for teaching, rules, etc.), the language pledge, or a combination of all these components that contributed faster RT in Spanish during such a short period of time. Recall that in Segalowitz and Freed (2004), learners were abroad for four months but there was no reference to a language pledge, and data were not collected *in situ*.

Speaking more broadly, these results open a new venue for comparison between the SA experience over the AH. Along these lines, in chapter 3 of this dissertation it was shown that there was a very obvious benefit of SA (i.e., our IM learner group) over AH based on production data, namely, in terms of oral fluency (e.g., Freed, 1995; García-Amaya, 2008, 2009, 2010; García-Amaya and Filgueras-Gómez, 2010; Díaz-Campos, 2006; Collentine, 2004; Lafford, 1995, 2004). In this section, we have evidence for SA learners that shows that in just six weeks of time there is a cognitive benefit for L2 processing and lexical access in the second language.

For now, a question that arises, given these findings, is what is the effect of the abroad experience for the same IM learners on processing of the L1. This is the focus of the next section.

4.3 Results of English SCT for accuracy and speed (English lexical access task)

In this section the results of the English version of the SCT are presented. As with the Spanish version of this task the English task also yielded accuracy and global speed results. This task replicated the design used by Segalowitz and Freed (2004). Also, instead of being conducted three times, like the Spanish SCT, this task was conducted twice, at the beginning and at the end of the program, with the intention of replicating the English SCT by Segalowitz and Freed (2004). Both the beginning and end versions of the English SCT were exactly the same and followed the same design and stimuli of task used by Segalowitz and Freed (2004), with the only difference that instead of six blocks of practice, participants were given ten.

4.3.1 Accuracy results

The accuracy analysis revealed that although the IM learners had much higher indexes of accuracy in the English SCT than in the Spanish SCT, they were also slightly less accurate at Time 2 than at Time 1. Table 4.3 shows the percentage for correct and incorrect responses at Times 1 and 2. A Generalized Linear Model was run to calculate the effect of TIME on a ACCURACY, a binary independent variable, and it showed a non significant effect, Wald Chi-Square = 3.740, p=.053. Also, there was no effect of PROFICIENCY on ACCURACY, Wald Chi-Square = .551, p=.458 or overall effect of the interaction of TIME and PROFICIENCY for ACCURACY, Wald Chi-Square = 4.105, p=.250.

Table 4.3. Accuracy English SCT

TASK	TIME 1	TIME 2
ACC (%)	96.6 %	95.5%

4.3.2 Global speed results

The global speed results of the English SCT are presented in Figure 4.7, which shows that the overall RT of IM learners for the English semantic classification task were statistically faster at Time 1 than at Time 2, F(1,5027) = 18, p=.000. This indicates that in six weeks, these participants' ability to access English words (i.e., their native language) was partially inhibited or attenuated.

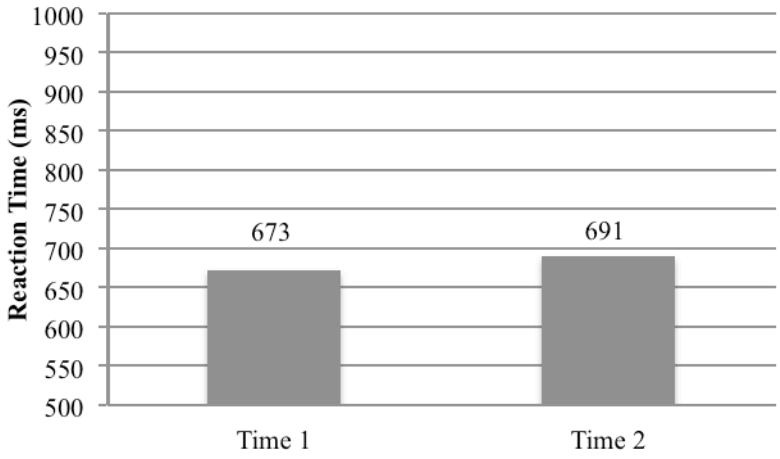

Figure 4.7. Mean RTs of English lexical access tasks for IM learners based on Time.

An additional analysis showed that there was a non-significant effect for PROFICIENCY on RT across TIME, F(1,25) = .560, p=.461, or either at Time 1, F(1,25) = 1.033, p=.319, or Time 2, F(1,25) = .253, p=.619. However, there was a significant effect for interaction of TIME and PROFICIENCY on RT for IM learners in the English SCT, F(1,5026) = 5.587, p=.018, which indicates that the two proficiency groups did not behave in similar ways across time. A detailed analysis of low and high proficiency IM learners follows.

4.3.3 Speed results for high and low proficiency learners

In this section, the data of the English SCT are presented according to levels of proficiency. Figure 4.8 shows the average RTs of low and high proficiency IM learners at Times

1 and 2. A Linear Mixed Model was run to analyze the effect of TIME on RT for low proficiency IM learners and showed that the effect of TIME was statistically significant, F(1,2586) = 22.74, p=.000, indicating that these subgroup of IM learners became slower accessing English words across time. In comparison to low proficiency IM learners, high proficiency IM learners showed very similar mean RT at Times 1 and 2. Moreover, the statistical analysis indicated that there was not a significant effect of TIME on RT for high proficiency IM learners, F(1,2435) = 1.49, p=.222.

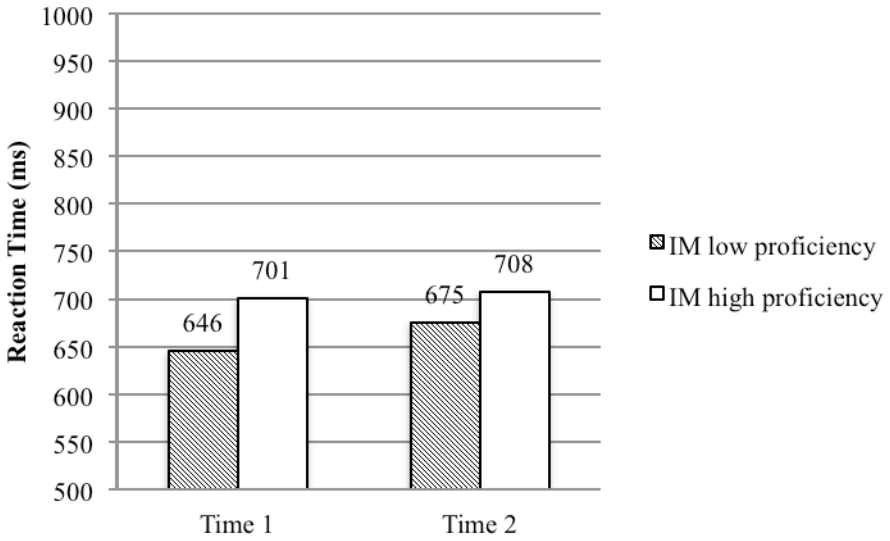

Figure 4.8. English SCT RT for low and high proficiency IM learners as a function of Time

The statistical differences for low and high proficiency are very important as they indicate that by the end of the program, high proficiency IM learners were accessing English words at the same rate as at the beginning, and low proficiency IM learners were also becoming statistically slower in their processing of English words.

4.3.4 Individual speed results for new items

Next, an individual analysis was conducted for each learner on correct responses only, and the results, presented on Figure 4.9, revealed that six of the 13 high proficiency IM learners had slower RTs at Time 2 than at Time 1, while seven had faster RTs.

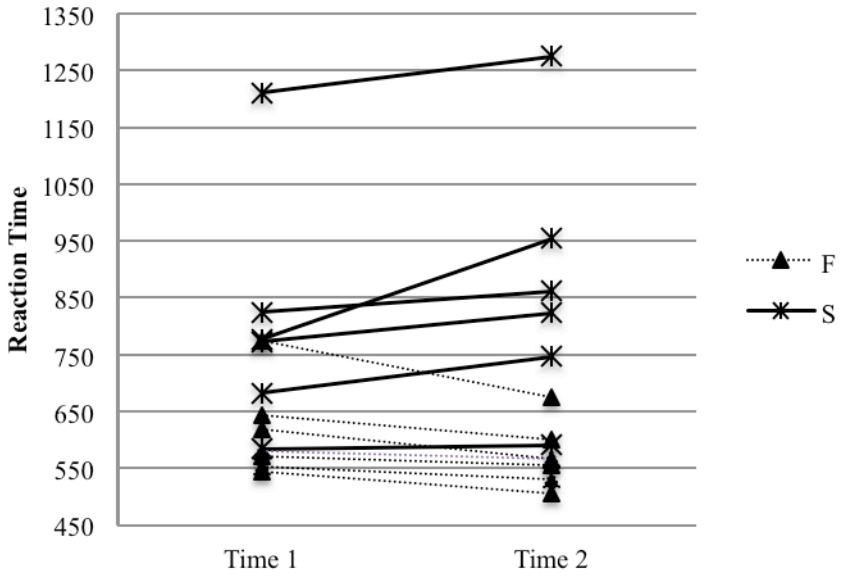

Figure 4.9. Individual results for the English SCT for high proficiency IM learners

Figure 4.10 shows the results for the English SCT for low proficiency learners and shows that nine of the 14 participants in this subgroup were slower at Time 2 than at Time 1, while five were faster at Time 2 than at Time 1. Low proficiency learners all remained within a similar range (i.e., there were no outliers).

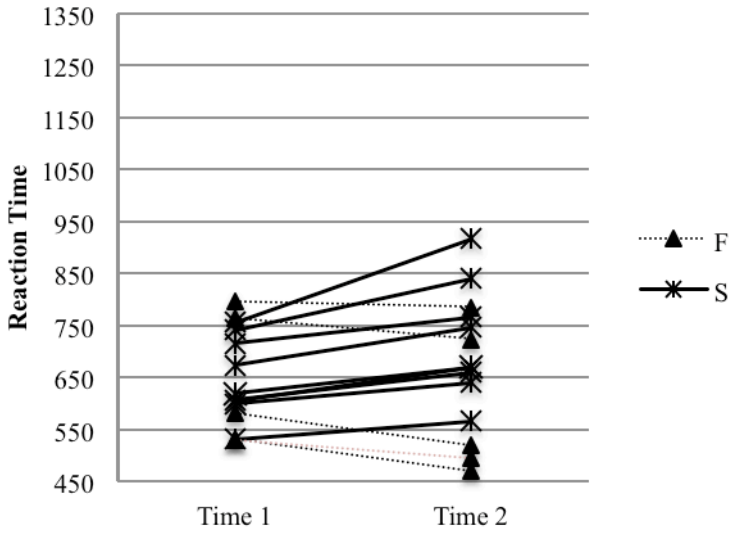

Figure 4.10. Individual results for the English SCT for low proficiency learners.

Finally, Table 4.4 shows the individual results of the English SCT. The table includes the specific RTs recorded at Time 1 and at Time 2 and if participants were slower (S) or faster (F) at

Time 2 than at Time 1. The data in the table also show the significant effect of TIME on RT for each learner. In total, 27 Linear Mixed Models were run, one per each participant included in the sample. The results indicated that out of the 16 participants that had slower RTs at Time 2, 12 of these were at the level of statistical significance. In the same fashion, 11 participants had faster RTs at Time 2 than at Time 1, and the statistical analysis indicated that seven of these learners were significantly faster.

Table 4.4. Individual results for the English SCT with statistical significance.

PROFICIENCY	PARTICIPANT	T1	T2	T1-T2	SIGNIFICANCE T1-T2
LOW	1	607	661	S*	F(1,171)=5.495, p=.020
LOW	2	763	723		F(1,186)=3.775, p=.054
LOW	4	716	767		F(1,177)=3.337, p=.069
LOW	8	795	785		F(1,181)=.113, p=.738
LOW	9	531	568	S***	F(1,187)=14.82, p=.000
LOW	11	742	841	S**	F(1,178)=9.460, p=.002
LOW	12	605	671	S***	F(1,185)=11.44, p=.001
LOW	13	532	469	F***	F(1,181)=34.57, p=.000
LOW	14	600	641	S*	F(1,176)=4.374, p=.038
LOW	17	756	919	S***	F(1,169)=26.29, p=.000
LOW	21	582	521	F***	F(1,177)=18.04, p=.000
LOW	23	531	495	F***	F(1,184)=12.69, p=.000
LOW	25	620	669	S*	F(1,178)=6.105, p=.014
LOW	27	673	747	S**	F(1,178)=6.904, p=.009
HIGH	3	774	674	F***	F(1,185)=28.43, p=.000
HIGH	5	777	954	S***	F(1,174)=33.96, p=.000
HIGH	6	772	823	S*	F(1,185)=5.474, p=.020
HIGH	7	618	565	F***	F(1,181)=13.68, p=.000
HIGH	10	571	555		F(1,176)=1.876, p=.173
HIGH	15	1210	1276		F(1,172)=1.276, p=.260
HIGH	16	581	567		F(1,184)=1.614, p=.206
HIGH	18	584	591		F(1,179)=.083, p=.773
HIGH	19	682	747	S***	F(1,187)=20.87, p=.000
HIGH	20	644	600	F***	F(1,182)=12.95, p=.000
HIGH	22	824	862		F(1,184)=1.429, p=.233
HIGH	24	544	506	F***	F(1,187)=11.64, p=.001
HIGH	26	553	530	S*	F(1,187)=5.618, p=.019

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

In sum, between Times 1 and 2, 12 participants became significantly slower, seven became significantly faster, and eight did not experience any significant development in their speed to access English words in their mental lexicon during the 6-week IM context.

4.3.5 Summary and discussion

Replication of the English SCT used in Segalowitz and Freed (2004) *in situ* with these IM learners proved to be fruitful endeavor. Certainly, it is not possible to ascertain if participants in Segalowitz and Freed (2004) would have performed similarly had they performed this task while they were still abroad. Recall that Segalowitz and Freed reported that after participating in a semester (i.e., four-month) SA program in Spain, the learners in their study had become faster at accessing words in their original version of the English SCT. It is surprising, then, that overall, participants in this dissertation became significantly slower at a similar version of the same task during only six weeks of immersion. These results, then, are indication that collecting data *in situ* might yield different results because lexical access might underlie different constraints depending on the L1 and L2 use factors. Certainly, these results do complement Segalowitz and Freed's data as it shows an earlier stage for L1 lexical access that had not been reported in the literature.

Another novel finding is the distribution of data between low and high levels of proficiency. Specifically, the results showed that the low proficiency group had a significantly slower RT at Time 2. The high proficiency IM group, on the contrary, maintained a constant RT for the English stimuli (even though there was an outlier that was particularly slow at both Times 1 and 2). This opposite trend based on proficiency may indicate that although high proficiency IM learners have inhibited their L1 while abroad, the level of inhibition did not change in a statistically significant fashion during the six weeks between the two data collection time points.

Moreover, the results are initial evidence that low proficiency learners might need to inhibit their native language more than high proficiency ones to adapt to the requirements of the immersion context. Indeed, the statistical analysis showed a significant interaction of TIME and PROFICIENCY on RT in the English lexical access task.

An individual analysis was carried out to shed light on this effect and revealed that within the high proficiency IM group, out of the 13 IM learners, six obtained slower RTs at Time 2, while the other seven obtained faster RTs over time. For low proficiency IM learners, it was seen that of the 14 learners in this group, nine were slower at Time 2 while five were faster. Possibly, the significance between Times 1 and 2 for this particular group was caused by the larger number of participants that slowed their RTs at Time 2. The individual analysis revealed that of the 27 participants in the sample, 12 became slower at Time 2, while seven became faster. This indicates that although originally, proficiency seemed to be the only reason behind the significance of the RT of low proficiency learners, not all learners behaved in the same way. Although a majority of them may have been influenced by either an inhibitory effect or an attenuation process in their L1 that caused them to become slower at Time 2, almost a third of the sample (i.e., seven participants) had a highly significant effect of TIME on their RT for this task but in the opposite direction by becoming faster, which indicates that they did not experience inhibition, at least not on the lexical access level, or to the extent that would make them slow down in accessing English words in the mental lexicon. Regardless of this, it is important to recall that there was a significant effect of TIME on RT for the stimuli in the English SCT as well as in the Spanish SCT, and overall, IM learners became slower with English stimuli and faster with Spanish stimuli.

More generally, such results are relevant for the study abroad literature (Freed, 1995;

Segalowitz and Freed, 2004; DeKeyser, 1986, García-Amaya, 2008, 2009, 2010; García-Amaya and Filgueras-Gómez, 2010; Díaz-Campos, 2006; Collentine, 2004; Lafford, 1995, 2004) and for applied psycholinguistics (Segalowitz and Freed, 2004; Sunderman and Kroll, 2009; Linck et al., 2009; Kroll, 1993; 1994; Kroll and Tokowicz, 2001; Kroll and Sunderman, 2003; Wang, Xue, Chen, Xue and Dong, 2007), as these findings have direct implications for modeling L2 speech production processes. Indeed, previous models of bilingual speech production have not considered the role that inhibition plays in a foreign language or either how the immersion context shapes both the development of the L2 as well as other linguistic processes in the L1 (e.g., Flege, 1987, DeBot, 1992). A handful of very recent studies have pointed to the fact that the L1 undergoes inhibitory/attenuating effects in language learning contexts where the first language is not used (e.g., Linck et al. 2009, Sunderman and Kroll, 2009), however none of these replicated the study (and therefore the results) of Segalowitz and Freed (2004). Sunderman and Kroll (2009) and Linck et al. (2009) compared learners with AH and SA experience but did not include a pretest/posttest design in their studies. Specifically, Sunderman and Kroll (2009) collected data in the US from learners that had and had not lived abroad, whereas Linck et al. (2009) collected data from their SA group three months into the SA experience. Certainly, the methodological improvement of the present study in which data were collected *in situ* at multiple points of the abroad experience is an aspect that has implications for future investigations in this area, such as the necessity of replicating the L1 lexical access task to investigate the role of L1 inhibition and its effect on development of the L2.

Given the present findings on lexical access, we focus in the next section on lexical retrieval in the L2 for the same population of learners.

4.4 Results of the Spanish PNT for accuracy and speed (Spanish lexical retrieval task)

4.4.1 Accuracy results

This section presents the accuracy and speed results for the Spanish PNT, the cognitive task intended to measure the lexical retrieval abilities of IM learners. Table 4.5 indicates IM accuracy rates over time. The accuracy results of the Spanish PNT show that IM learners named correctly 75.1% of the stimuli presented, while they named incorrectly (or simply could not name) 24.9% of the stimuli. However, when the RTs were separated between NEW and OLD stimuli, the results indicate that while the accuracy difference between NEW and OLD stimuli was very small at Time 2, this difference increased at Time 3. The developmental pattern of the accuracy rates for PNT has a 'bell-shaped' form, as it increases at Time 2 and then goes down to the same level of Time 1. In order to analyze the effect of TIME on ACCURACY, a Generalized Linear Model was conducted. The analysis showed that for stimuli (i.e., NEW and OLD), there was a significant effect of TIME, between Times 1 and 2, Wald Chi-Square = 29.886, p=.000, and Times 1 and 3, Wald Chi-Square = 29.886, p=.000, but not between Times 2 and 3, Wald Chi-Square = 29.886, p=.788. A second model was run for NEW data only and the results indicated that there was an effect of TIME on ACCURACY between Time 1 and 2, Wald Chi-Square = 16.734, p=.002; and also between Times 2 and 3, Wald Chi-Square = 29.886, p=.000, but not between Times 1 and 3, Wald Chi-Square = 29.886, p=.795. Thus, the difference in accuracy between the beginning and end of the program is basically the same. A third model analyzed the effect of TIME on ACCURACY for OLD stimuli only and it revealed a significant effect of TIME, Wald Chi-Square = 19.577, p=.000, which indicated that IM learners were more accurate with OLD stimuli at Time 3 than at Time 2. Finally, the effect of NEW vs. OLD on ACCURACY was also analyzed and a significant effect was found at Time 2, Wald Chi-Square

= 14.499, p=.000, as well as Time 3, Wald Chi-Square = 167.550, p=.000, which indicates that at both times IM learners were significantly more accurate with OLD stimuli than with NEW ones.

Table 4.5. Mean accuracy for the IM group for the Spanish PNT.

TASK	TIME 1	TIME 2		TIME 3	
ACC (%)	69.2%	77.7%		78.1%	
STIMULI	NEW	NEW	OLD	NEW	OLD
ACC (%)	69.2%	75.4%	79.8%	69.7%	85.6%

4.4.2 Global speed results

This section introduces the results for the RTs obtained by IM learners for the PNT for NEW stimuli and correct responses only. Figure 4.11 shows that overall, IM learners obtained very similar RTs across time in the lexical retrieval task.

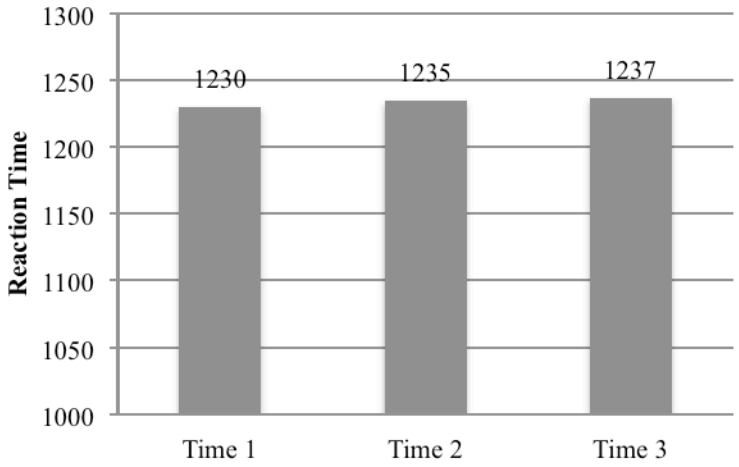

Figure 4.11. RTs for Spanish PNT for IM learners based on time (correct responses and new words only).

A Linear Mixed Model was run to calculate the effect of TIME on RT and revealed that that there was no effect of TIME on RT between Times 1 and 2, F(2,3888) = .115, p=1.000, between Times 1 and 3, F(2,3888) = .115, p = 1.000, or between Times 2 and 3, F(2,3888) = .115, p = 1.000), indicating that IM learners did not significantly increase or decrease their efficiency of lexical retrieval during the 6-week IM program. The next section presents the RT data separated according to levels of proficiency.

4.4.3 Speed results for high and low proficiency learners

This section presents the results of RT for the PNT for NEW stimuli and correct responses only for IM learners according to levels of proficiency. The statistical analysis revealed that there was no effect of PROFICIENCY on RT, $F(1,25) = .031$, $p=.862$, and no effect of the interaction of TIME and PROFICIENCY on RT, $F(2,3888) = 1.910$, $p=.148$. As Figure 4.12 shows, the RT data indicate that high and low proficiency IM learners had very similar RTs across time. The statistical analysis showed that there was no effect of TIME on RT for either high proficiency learners at any time (i.e., Times 1 and 2, $F(2,1937) = .483$, $p=.978$; Times 1 and 3, $F(2,1937) = .483$, $p=1.000$; or Times 2 and 3, $F(2,1937) = .483$, $p=1.000$), or low proficiency learners (i.e., Times 1 and 2, $F(2,1942) = 1.181$, $p = .425$; Times 1 and 3, $F(2,1942) = 1.181$, $p = 1.000$; and Times 2 or 3, $F(2,1942) = 1.181$, $p=.679$).

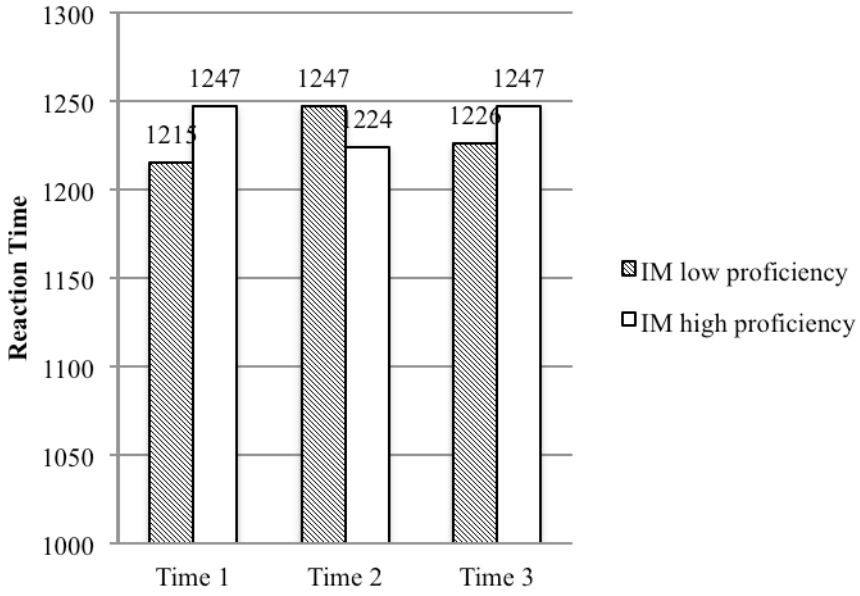

Figure 4.12. Spanish PNT average RT for low and high proficiency IM learners across time

The data show that there was no PROFICIENCY effect on RT and that the IM learners overall did not show any significant change in terms of lexical retrieval across time. The next section analyzes the RT differences for NEW and OLD stimuli.

4.4.4 Speed results for new and old stimuli

This section presents the results of overall RT for all IM learners for NEW stimuli at Time 1 and NEW and OLD stimuli at Times 2 and 3 according to levels of proficiency. Results for high proficiency learners are shown in Figure 4.13.

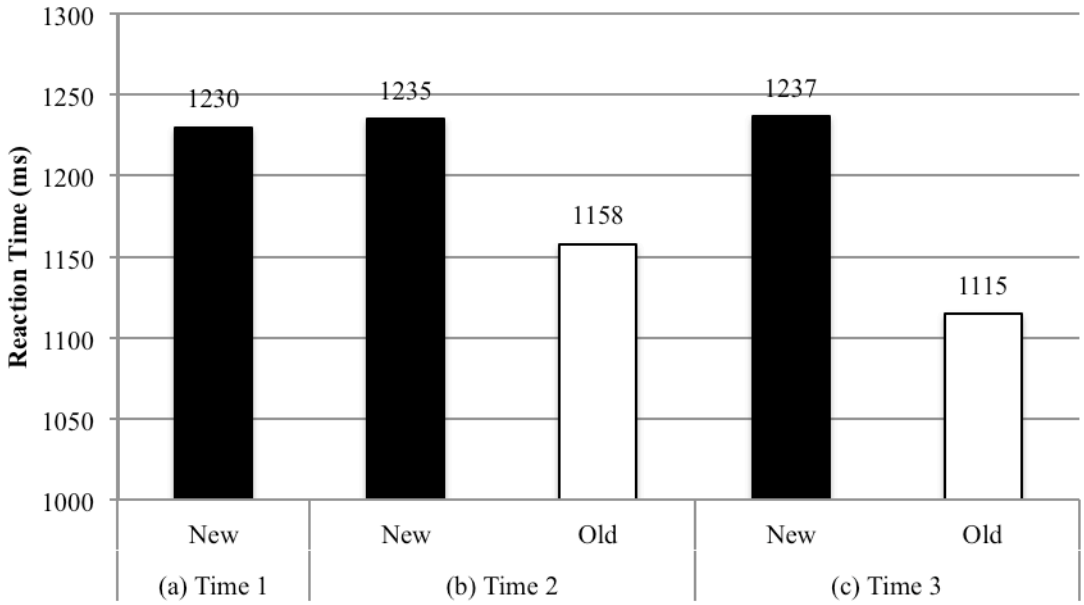

Figure 4.13. Mean RTs for Spanish PNT for NEW and OLD stimuli for all IM learners based on TIME

The effect of TIME on RTs for and OLD stimuli was statistically significant at Time 2, $F(1,2277) = 38.599$, $p=.000$) as well as at Time 3, $F(1,2272) = 87.900$, $p=.000$. However, these differences in OLD stimuli are most likely due to a practice effect, and are not of major interest in this case.

4.4.5 Speed results for new and old stimuli for high and low proficiency learners

This section presents the RT results for NEW and OLD stimuli in the PNT for high proficiency learners. The statistical analysis shows for low proficiency learners, there was a significant effect of NEW vs. OLD on RT at Time 2, $F(1,1137) = 15.216$, $p=.000$, as well as at Time 3, $F(1,1150) = 64.351$, $p=.000$. There was also a significant effect of NEW vs. OLD on RT

at Times 2, F(1,1133) = 24.083, p = .000, and at Time 3, F(1,1115) = 27.293, p = .000**.** So far, these data show that the statistical differences on RT between NEW and OLD stimuli also hold true when the IM learners are separated according to levels of proficiency. Restricting the analysis to old items only, we find that there was no reaction time difference for high proficiency learners, F(1,1284) = 3.027, p=.082, but a significant decrease in RT at time 3 for low proficiency learners F(1,1294) = 5.644, p=.018. Overall, these data show that there is no effect of PROFICIENCY, TIME, or NEW vs. OLD stimuli, on RT. This implies that lexical retrieval for NEW items did not improve. The next section introduces the individual analysis for the Spanish lexical retrieval task (i.e., PNT task).

4.4.6 Individual speed results for new items

This section presents the individual data of the PNT separated according to levels of proficiency. Figure 4.14 shows that depending on the pattern of speed for new items (correct responses only) there were seven possible trends that IM learners followed. This organization of trends follows the same structure that was used to describe the individual data for the Spanish SCT. The samples included in the legend are three letter combinations for which the first letter stands for the difference between Time 2 and Time 1 (i.e., F for faster and S for slower), the second letter stands for the difference between Times 2 and 3, and the third letter stands for the overall difference between Times 1 and 3.

Figure 4.15 indicates that the distribution of trends for the 13 high proficiency participants was the following: 2 F-F-F, 1 F-S-F, 2 F-S-S, 2 S-S-S, 1 S-F-S, and 5 S-F-F. When the difference between the beginning and end of the program was analyzed in isolation, it can be seen that eight participants were faster at the end, and five were slower at the end.

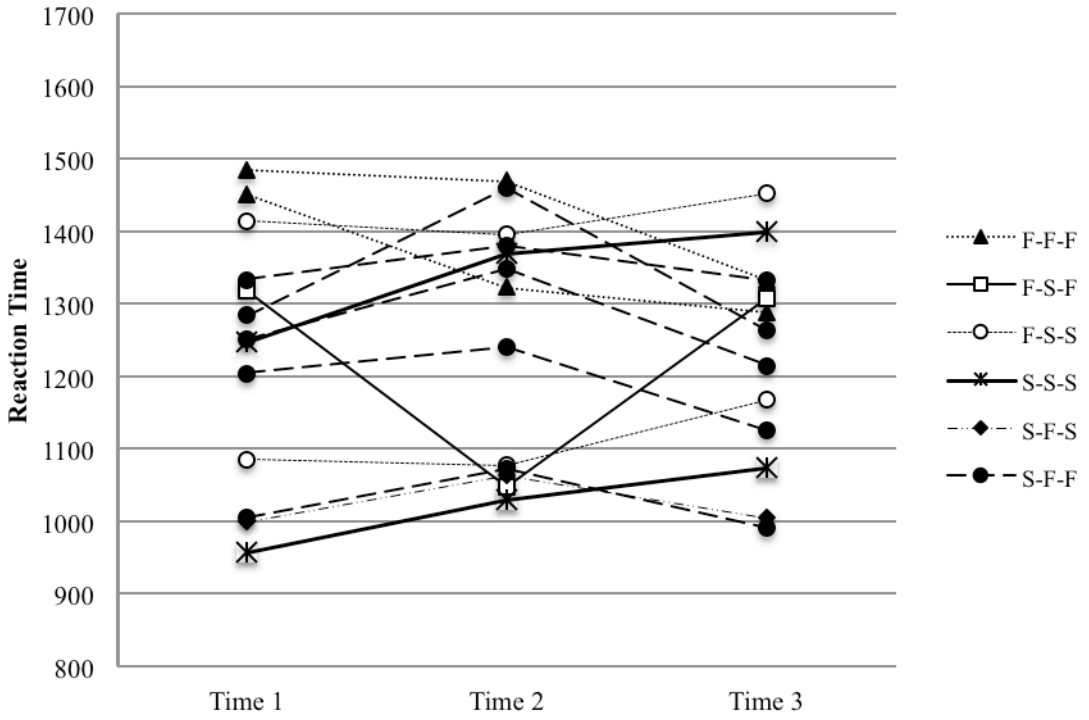

Figure 4.14. Spanish PNT RT individual IM learner difference for high proficiency learners

Figure 4.15 presents the individual patterns for low proficiency learners. This figure contains the same legend used in Figure 4.14 and indicates that out of the 14 low proficiency IM learners in the sample there were 3 F-F-F learners, 2 F-S-F learners, 3 F-S-S learners, 2 S-S-S learners, 2 S-F-S learners, and 2 S-F-F learners. When the trends between Times 1 and 3 are analyzed separately, it is seen that while seven low proficiency learners were faster between the beginning and the end of the program, six were slower.

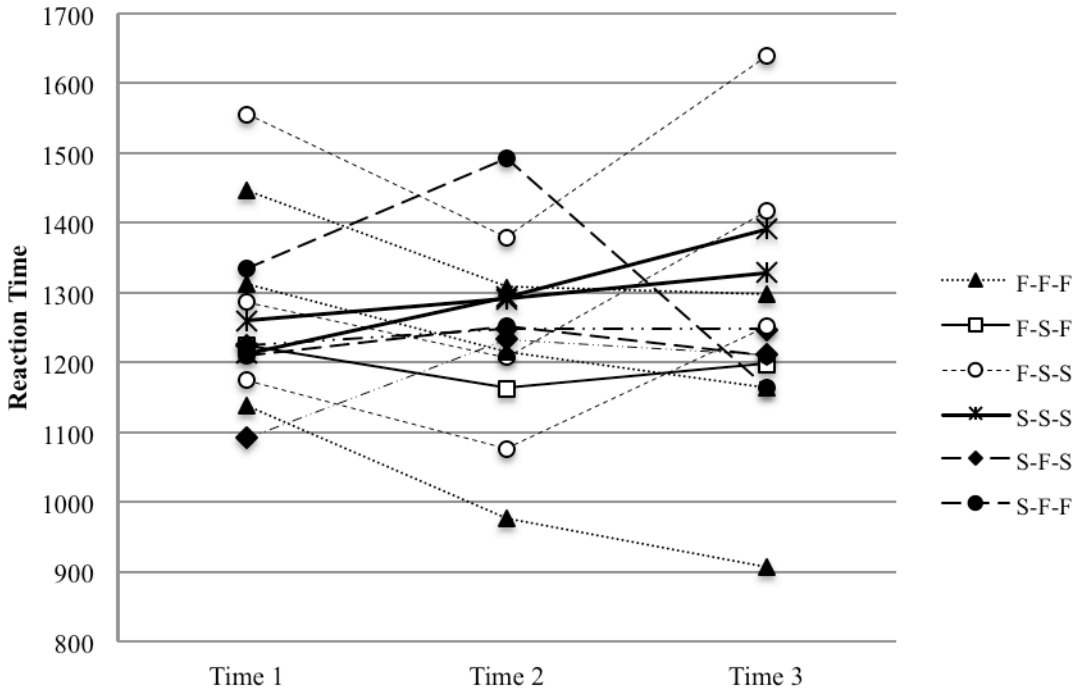

Figure 4.15. Spanish PNT RT individual IM learner difference for low proficiency learners

Table 4.6 presents the individual analysis for the Spanish PNT. When all data for the PNT were analyzed individually, the results showed that 13 participants were faster and 14 were slower at Time 2 than at Time 1. Of the 13 faster participants, two that were faster showed a statistically signficant result. Of the 14 participants that were slower at Time 2, two had a statitically significant result. When comparing the last three weeks of the program (i.e., results from Times 2 and 3), 15 participants were faster and 12 were slower. Of these participants 2 had a statistitically significant faster RT at Time 3, and 3 were statistically slower at Time 3. Finally, the individual difference for the PNT is analyzed from the beginning to the end of the program, it can be seen that of 15 participants that are faster at Time 3 than at Time 1 only two have speed RT that are statistically signficant, and that of 12 participants that are slower at Time 3 than at Time 1, only 1 has a RT that is statistically significant.

Table 4.6. Spanish PNT RT individual IM learner difference for low and high proficiency learners with significance.

PROF.	PART.	T1	T2	T3	T1-T2	T2-T3	T1-T3	SIGNIFICANCE T1-T2	SIGNIFICANCE T2-T3	SIGNIFICANCE T1-T3
LOW	1	1224	1162	1198				F(1,116) = 1.79, p=.184	F(1,70) = .167, p=.684	F(1,109) = .202, p=.654
LOW	2	1286	1207	1418		S*		F(1,86) = 1,89, p=.172	F(1,54) = 4.64, p=.036	F(1,89) = 2.581, p=.112
LOW	4	1092	1234	1211				F(1,101) = 2.388, p=.125	F(1,53) = 1.262, p=.266	F(1,97) = 1.902, p=.171
LOW	8	1214	1294	1393			S**	F(1,105) = 3.397, p=.068	F(1,62) = .729, p=.396	F(1,98) = 7.733, p=.007
LOW	9	1260	1291	1329				F(1.69) = .031, p=.862	F(1,39) = .253, p=.618	F(1,66) = .889, p=.349
LOW	11	1555	1379	1639		S*		F(1,79) = 2.592, p=.111	F(1,51) = 5.68, p=.021	F(1,77) = 1.424, p=.236
LOW	12	1335	1492	1164	S*	F***	F**	F(1,88) = 4.745, p=.032	F(1,66) = 27.464, p=.000	F(1,97) = 7.664, p=.007
LOW	13	1137	1078	1130				F(1,104) = .334, p=.564	F(1,67) = 1.853, p=.178	F(1,103) = .100, p=.753
LOW	14	1210	1253	1209				F(1,97) = .001, p=.972	F(1,66) = .033, p=.856	F(1,95) = .071, p=.791
LOW	17	1446	1308	1297				F(1,85) = 2.257, p=.137	F(1,61) = .027, p=.870	F(1,81) = 3.031, p=.085
LOW	21	1313	1216	1164				F(1,90) = .919, p=.340	F(1,66) = .918, p=.342	F(1,85) = 3.721, p=.057
LOW	23	1174	1077	1253		S*		F(1,95) = 3.037, p=.085	F(1,66) = 4.67, p=.034	F(1,82) = 1.008, p=.318
LOW	25	1137	976	906	F**		F***	F(1,101) = 9.275, p=.003	F(1,53) = .123, p=.728	F(1,99) = 11.958, p=.001
LOW	27	1225	1248	1247				F(1,93) = .056, p=.813	F(1,64) = .026, p=.873	F(1,93) = .355, p=.553
HIGH	3	1413	1396	1452				F(1,105) = .877, p=.351	F(1,69) = .241, p=.625	F(1,107) = .009, p=.924
HIGH	5	1451	1323	1289	F*			F(1,95) = 6.224, p=.014	F(1,53) = .002, p=.962	F(1,89) = 2.465, p=.120
HIGH	6	1247	1370	1400				F(1,103) = 1.890, p=.172	F(1,61) = .099, p=.755	F(1,106) = 2.653, p=.106
HIGH	7	957	1030	1074			S*	F(1,97) = 2.012, p=.159	F(1,50) = 2.241, p=.141	F(1,94) = 4.666, p=.033
HIGH	10	999	1064	1005				F(1,124) = 1.338, p=.250	F(1,63) = .668, p=.417	F(1,119) = .035, p=.852
HIGH	15	1485	1470	1332				F(1,73) = .031, p=.861	F(1,53) = .599, p=.442	F(1,68) = 1.612, p=.209
HIGH	16	1085	1077	1167				F(1,111) = .020, p=.887	F(1,63) = 2.161, p=.147	F(1,106) = 2.051, p=.155
HIGH	18	1250	1347	1214				F(1,103) = .862, p=.355	F(1,65) = 1.58, p=.213	F(1,91) = .547, p=.462
HIGH	19	1204	1240	1125				F(1,101) = .052, p=.821	F(1,64) = .527, p=.470	F(1,95) = .733, p=.394
HIGH	20	1006	1073	991		F*		F(1,126) = 1.774, p=.185	F(1,79) = 4.023, p=.048	F(1,125) = .311, p=.578
HIGH	22	1318	1047	1307	F***	S***		F(1,113) = 18.734, p=.000	F(1,73) = 22.642, p=.000	F(1,103) = .003, p=.954
HIGH	24	1284	1460	1263	S*	F*		F(1,98) = 6.98, p=.010	F(1,68) = 4.102, p=.047	F(1,94) = .012, p=.914
HIGH	26	1333	1379	1332				F(1,89) = .772, p=.382	F(1,56) = .093, p=.774	F(1,82) = .002, p=.964

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

4.4.7 Summary and discussion

The Spanish PNT used in this dissertation was designed to examine lexical retrieval development in six weeks in an intensive overseas immersion program. The retrieval process of words does not only require accessing the word but also articulating it (i.e., according to the results in this dissertation there is approximately 600 ms difference between learner accessing words and retrieving them).[17] One of the aspects taken into account in the design of the task was the possible effect that remembering the stimuli may have on the subsequent versions of the task (i.e., at Times 1 and 2). Thus, the combination of all new stimuli at Time 1, 50% new and 50% old stimuli at Time 2, and 50% new and 50% old stimuli (i.e., 25% from Time 1 and 25% from Time 2) at Time 3 was very informative, as it showed that as hypothesized, IM learners were always faster with old stimuli than with new stimuli. This supported our initial hypothesis that by having three data collecting time points over six weeks, participants would become faster processing stimuli that have been shown to them in previous versions of the task (i.e., at Times 1 and 2). In the same fashion, this finding shows that failure to include new stimuli would have led us to infer that IM learners experienced a significant reduction in lexical retrieval latencies as well as that IM learners significantly improved the efficiency with which they retrieve words in their L2.

Accuracy for all stimuli (i.e., NEW and OLD) increased across time, particularly during the first three weeks (i.e., between Times 1 and 2) and also across time during the entire program (i.e., between Times 1 and 3), however there was not a significant change on ACCURACY during the last three weeks of the program (i.e., between Times 2 and 3). When ACCURACY

[17] It in the task would provoke learners to retrieve the Spanish version of the word, others would cause them to retrieve an was also noted that while some of the pictures used English word instead, in most cases they would retrieve the first syllable in English and then, when they knew the Spanish word, correct themselves. Overall, future research investigating similar learners as well as others from different levels may be counterbalanced with other cognitive tasks such as translation tasks.

was analyzed for NEW stimuli only, similar patters were found, and it was only during the last three weeks of the program when an effect of TIME on ACCURACY was not seen. Also and as expected, IM learners were more accurate with OLD stimuli at Time 3 than at Time 2. This was a reasonable hypothesis, as at Time 3, IM learners would have had more experience with OLD stimuli.

The results obtained in the Spanish PNT indicate that, overall, participants were not able to significantly reduce their lexical retrieval latencies for new stimuli during the six-week program. There were no differences when learners were separated according to levels of proficiency, and IM learners only showed statistically significant improvement with OLD stimuli. On the other hand, this result contrasts with the developmental results of the Spanish lexical access task (i.e., SCT), where significant changes were seen across time for new stimuli for low and high IM learners. Indeed, the overall RTs for OLD stimuli were significantly faster at Time 3 than at Time 2. However, once the data were separated according to levels of proficiency, a statistically significant effect for TIME on RT was found for low proficiency learners only. Follow-up analysis of individual data showed that two speakers had significant faster speed between Times 1 and 3. In the same way, only one of participants was significantly slower between the beginning and end of the program, indicating that most participants did not show a significant increase in lexical retrieval efficiency across time.

As pointed out in Chapter 1, a similar picture-naming task was used by Sunderman and Kroll (2009). Data from the picture-naming task used in their study were collected for two groups of learners, one with and one without experience abroad, in the US. Their data show that learners with SA experience were faster and more accurate than the learners without SA experience in comprehension and production. The authors mentioned further that there was a

partially significant interaction between SA experience and working memory, one of the tenets of their study. In addition, the authors reported on support for the external cue hypothesis (faster and more accurate L2 processing in the L2 environment due to salient information in the SA context), which suggests that learners with higher internal resources would benefit more from the SA context (also see Tokowicz et al., 2004). Notwithstanding the differences between Sunderman and Kroll's picture-naming task and the one used in this dissertation, the data obtained for our IM learners in which there is no significant change (at least for new words) over six weeks would not initially support the results obtained by Sunderman and Kroll for the learners with experience abroad for speed of lexical retrieval. However, it is possible that this result is due to a difference in the amount of time that learners in both studies spent abroad, as the SA learners in the Sunderman and Kroll study spent an average 3.8 months abroad, whereas the IM learners in this dissertation spent six weeks only abroad. Given this discrepancy, there is a possibility that time of exposure in the L2 had long term effects on the learners in Sunderman and Kroll's study.

4.5 Results of the English ACT for accuracy and speed (English Attention Control Task)

In order to analyze the ACT it is necessary to compare shift trial (CONDITION1 and CONDITION3) with the baseline – or no-shift – trial (CONDITION0). CONDITION1 and CONDITION3 behave in the same way but they are slightly different, because in CONDITION3, a correct answer could be due to an answer perseveration. However, an error is unambiguous. The most important aspect of CONDITION3 therefore is error rate. Due to this slight difference between both conditions, it is necessary to take either CONDITION1 (or CONDITION3) and compare it with CONDITION0. The group analysis shows that ACCURACY is higher, and RTs are faster, in CONDITION0 compared to the shift CONDITION1, across times, as shown in Table 4.7. Then, it is

also important to compare if CONDITION1 and CONDITION3 differ significantly to verify the possible presence of an answer perseveration in CONDITION3. It is possible to collapse CONDITION1 and CONDITION3 if there are no significant differences between the two. If there were a significant difference (e.g., higher accuracy CONDITION3) it could be due to perseveration. If there is no significant difference between 1 and 3 that means that answer perseveration was not an important factor in the accuracy performance for CONDITION3. CONDITION1 is clear in the sense that a high accuracy rate is unambiguously indicating an ability to shift attention. A Generalized Linear Model was run to analyze ACCURACY, as this statistical procedure allows for a binary dependent variable.

Table 4.7. Accuracy rate for CONDITION0, CONDITION1 and CONDITION3

		Across Time	Time 1	Time 2
ACCURACY RATE	CONDITION 0	86.1%	86.1 %	86.1%
	CONDITION 1	79.4%	79.8 %	79 %
	CONDITION 3	77.2%	77 %	77.4 %

For CONDITION0, the accuracy rate is exactly the same across time, (i.e., at Times 1 and 2). Also for CONDITION1, accuracy is 79.4% across Time, 79.8% at Time 1 and 79% at Time 2. For CONDITION0 and CONDITION1, there is a highly significant effect of CONDITION on ACCURACY across time, Wald Chi-Square = 37.695, p=.000, showing that listeners were more accurate in the "no-shift" condition than in the "shift" condition. There is also a significant effect on ACCURACY at TIME 1 only, Wald Chi-Square = 19.609, p=.000, and also at TIME 2 only, Wald Chi-Square = 20.715, p=.000.

For CONDITION0 and CONDITION3, there is a highly significant effect of CONDITION on ACCURACY across time, Wald Chi-Square = 62.376, p=.000, showing that here also, listeners

were more accurate in the "no-shift" condition than in the "shift" condition. There is also a significant effect on ACCURACY for TIME 1 only, Wald Chi-Square = 26.191, p=.000, and also for TIME 2 only, Wald Chi-Square = 33.835, p=.000. For CONDITION1 and CONDITION3, as expected, there is no significant effect of CONDITION on ACCURACY, Wald Chi-Square = 2.320, p=.128. Moreover, there was no effect of TIME on ACCURACY, Wald Chi-Square = .028, p=.867, suggesting that the group performance did not evolve in any significant way over the course of 6 weeks.

Table 4.8 presents the mean RT for correct responses only for CONDITION0 and CONDITION1 and CONDITION3. The analysis of RTs reveals that for CONDITION0 and CONDITION1, the effect of CONDITION on RT is highly significant across time, F(1,5462) = 43.144, p=.000; at Time 1 only, F(1,2720) = 20.846 p=.000; and at Time 2 only, F(1,2715) = 24.582, p=.000, indicating that at each instance IM learners were significantly faster with CONDITION0 than with CONDITION1. This parallels the accuracy data. Interestingly however, for these two conditions, the statistical analysis also unveiled a significant effect of TIME on RT for CONDITION0, F(1,4434) = 26.517, p=.000, and CONDITION1, F(1,1001) = 6.896, p=.009, indicating that IM learners were significantly faster at the end of the program (i.e., the second time they performed the task after a total of six weeks), than at the beginning, and this happened for both conditions (shift and no-shift) equally.

Table 4.8. Mean RT for correct responses only for CONDITION0 and CONDITION1 and CONDITION3

		Across Time	Time 1	Time 2
MEAN RT (ms)	CONDITION 0	1016	1043	988
	CONDITION 1	1097	1126	1067
	CONDITION 3	1073	1109	1037

A statistical analysis comparing RT differences for CONDITION0 and CONDITION3 was not run for two reasons. First, since CONDITION3 is mostly interesting in terms of error rates, and since RT are computed over correct answers only, it might not be straightforward to analyze the findings. Second, computing RT for erroneous answers only is also complicated by the fact that it is an error. While this could possibly be done, of course, we did not expect a different pattern from the comparison obtained for CONDITION0 and CONDITION1. As our analysis showed, there was no difference in accuracy or in RT between CONDITIONS 1 and 3. Therefore, the pattern we observe for CONDITION0 and CONDITION1 is likely representative of the learners' performance in this task.

In addition, an individual analysis of the accuracy for CONDITION0 and CONDITION1 was also conducted. Table 4.9 shows the individual results presented according to level of proficiency. A series of Generalized Linear Models were run to analyze the effect of CONDITION on ACCURACY, and the results indicated that there was a significant effect for all participants. However, 25 of them were in the direction of being more accurate with CONDITION0 than CONDITION1 while three low proficiency learners and one high proficiency learner (marked in bold on Table 4.9) were more accurate on CONDITION1 than CONDITION0. In other words, the majority of IM learners, with the only exception of four participants, were significantly more accurate in the "no-shift" condition than in the "shift" condition.

Table 4.9. English ACT accuracy individual IM learner difference for low and high proficiency learners with significance

PROF.	PART.	COND. 0 (ACC)	COND. 1 (ACC)	EFFECT COND. ON ACC. (Wald Chi-Square and p-value)	COND. 0 (RT)	COND. 1 (RT)	EFFECT COND. ON RT. (F-value and p-value)
LOW	1	**79.20%**	**83.30%**	0.00, p=.000	1092	1127	F(1,188) = .104, p=.748
LOW	2	83.90%	77.10%	0.00, p=.000	1139	1279	F(1,194) = 5.665, p=.018
LOW	4	82.80%	72.90%	0.00, p=.000	987	1058	F(1,190) = 1.414, p=.236
LOW	8	89.10%	87.50%	0.00, p=.000	1241	1369	F(1,210) = 3.463, p=.064
LOW	9	89.60%	85.40%	0.00, p=.000	952	1071	F(1,210) = 4.073, p=.045
LOW	11	**88.00%**	**89.60%**	0.00, p=.000	948	1062	F(1,209) = 2.447, p=.119
LOW	12	94.80%	83.30%	0.00, p=.000	1068	1125	F(1,219) = 1.047, p=.307
LOW	13	**81.80%**	**83.30%**	0.00, p=.000	1082	1103	F(1,194) = .005, p=.944
LOW	14	61.50%	47.90%	0.00, p=.000	1053	1060	F(1,138) = .002, p=.961
LOW	17	72.90%	68.80%	0.00, p=.000	751	711	F(1,170) = .618, p=.433
LOW	21	90.10%	79.20%	0.00, p=.000	1069	1103	F(1,208) = .242, p=.623
LOW	23	95.80%	85.40%	0.00, p=.000	860	1026	F(1,222) = 7.784, p=.006
LOW	25	92.70%	91.70%	0.00, p=.000	1049	1221	F(1,219) = 6.699, p=.010
LOW	27	87.00%	79.20%	0.00, p=.000	1118	1203	F(1,202) = 1.582, p=.210
HIGH	3	**91.10%**	**91.70%**	0.00, p=.000	946	940	F(1,215) = .186, p=.667
HIGH	5	84.50%	75.00%	0.00, p=.000	1094	1205	F(1,195) = 3.105, p=.080
HIGH	6	67.20%	58.30%	0.00, p=.000	1312	1392	F(1,153) = 2.529, p=.114
HIGH	7	83.30%	81.30%	0.00, p=.000	869	945	F(1,196) = 2.403, p=.123
HIGH	10	87.00%	81.30%	0.00, p=.000	917	983	F(1,203) = 1.255, p=.264
HIGH	15	88.50%	83.30%	0.00, p=.000	1384	1353	F(1,207) = .177, p=.674
HIGH	16	90.60%	83.30%	0.00, p=.000	966	1081	F(1,211) = 4.833, p=.029
HIGH	18	91.10%	72.90%	0.00, p=.000	841	932	F(1,207) = 2.425, p=.121
HIGH	19	90.60%	79.20%	0.00, p=.000	993	1086	F(1,209) = 1.801, p=.181
HIGH	20	94.80%	75.00%	0.00, p=.000	922	945	F(1,215) = .322, p=.571
HIGH	22	88.00%	83.30%	0.00, p=.000	1109	1261	F(1,206) = 6.315, p=.013
HIGH	24	83.90%	77.10%	0.00, p=.000	883	1020	F(1,195) = 4.102, p=.044
HIGH	26	93.80%	87.50%	0.00, p=.000	862	961	F(1,219) = 3.968, p=.048

With regard to effect of CONDITION on RT, a series of Linear Mixed Models were run, and the results indicated that there was a significant effect of CONDITION for a total of eight participants (marked in grey on Table 4.9): four low proficiency learners and four high proficiency learners.

4.6. Summary and discussion

The ACT used in this dissertation was originally designed by Darcy et al. (2011), who

used it in a cross-sectional design. This dissertation took it one step forward and analyzed this task from a longitudinal perspective. As was explained in Chapter 2, the English ACT task was modeled around five different conditions through which listeners, in this particular case, IM learners, would be tested on their efficiency to change levels of attention through a set of stimuli organized to require the use of attention control. More specifically, the main goal of this task was to account for accuracy and speed of processing. Three conditions were analyzed in this dissertation, the "no-shift" conditions (i.e., CONDITION0), and the two "shift" conditions (i.e., (CONDITION1 and CONDITION3). First, CONDITION0 is the no-shift between levels and thus it was expected that participants would reach high levels of accuracy. CONDITION1 and CONDITION3 necessitate an ability to shift attention to different information in the signal (voice or lexical status) in order to make an accurate answer. RT can therefore be slower, revealing a difficulty with shifting attention. As well, accuracy can be lowered by a difficulty to shift attention.

As expected IM learners were significantly more accurate on CONDITION0 than CONDITION1 and CONDITION3. Also, there were no significant differences in accuracy between CONDITION1 and CONDITION3, which indicated that IM learners were more accurate in the "no-shift" than in the "shift" conditions. Then, the individual analysis for efficiency showed that of 27 participants, four only were more accurate with the "shift" condition than with the "no-shift" condition. As for speed of processing, the RT results showed that participants were also faster with CONDITION0 than with CONDITION1, which also showed that IM learners were faster with the "no-shift" than with the "shift" condition. The individual analysis for speed of processing also showed that nine participants (a third of the sample size) had significant faster RT for CONDITION0 than CONDITION1, while the rest of the learners showed no significant differences.

As we have seen, the analysis of this task was carried out by comparing accuracy and/or

RT among the three most revealing conditions. This is a different analysis from the one that was carried out for the other cognitive tasks where the most important factor was the development of processing speed across time. The hypothesis of using a task in which change for individual conditions across time was not expected justifies its use, as a way of complementing this investigation, which already included a number of cognitive tasks where change across time was expected (i.e., Spanish and English SCTs, and Spanish PNT – although no change was observed for the latter). As was expected, the accuracy of the English ACT shows that there is no change for individual conditions across time. Therefore, the English ACT justifies the changes found in the other tasks and shows a cognitive ability tested in the L1 that was not affected by the L2 immersion context.

4.7 General discussion and conclusion

4.7.1 Overall trends

The four cognitive tasks analyzed in this chapter offer new insight into the way mental processes develop for IM learners during a complete L2 immersion where there is an institutionalized pledge to avoid the use of the L1. The results obtained through four cognitive tasks (i.e., Spanish SCT, English SCT, Spanish PNT and English ACT) were analyzed considering two levels of proficiency for each group, and individual differences were also reported. Before focusing on the individual variation, it is necessary to summarize the trends for high and low proficiency learners. In Table 4.10, the arrows and flat lines represent the different performance of the proficiency subgroups: significantly higher RT (upward arrow), or significantly lower RT (downward arrow), and no significant change (horizontal line).

Table 4.10. Proficiency groups trends on Spanish SCT, English SCT and Spanish PNT

	Spanish SCT	English SCT	Spanish PNT
Low proficiency IM learners			
High proficiency IM learners			

As the table shows, high proficiency learners overall became statistically faster at accessing words in Spanish in their mental lexicon, but showed no significant changes in the two other tasks (i.e., English SCT and Spanish PNT). On the other hand, low proficiency learners exhibited significantly faster RT for both the Spanish SCT. Low proficiency IM learners were also the only learners to become significantly slower at any of the cognitive tasks included in this dissertation (i.e., English SCT). As for the Spanish PNT, low proficiency IM learners behaved in a way similar to the high proficiency IM learners and did not show any significant change in their speech across time.

Focusing on the English SCT and more specifically on the performance of low proficiency IM learners, the two sets of data would indicate that overall, low proficiency IM learners followed completely different trends (i.e., overall significantly slower results for the lexical access task at Time 2). Considering the particular learning context in which the IM learners were immersed, these data could be interpreted from an inhibitory/L1 attenuation perspective. We focus on this issue in more depth in the following section.

4.7.2 Main highlights

As was pointed out at the beginning of this chapter, the IM learners only performed the cognitive tasks because they were the learners that showed significant gains in fluency measures in their immersion-learning context. Also, it was important to investigate if changes at the cognitive level would take place during the six-week period of immersion. The investigation of cognitive abilities revealed very important results, showing that Spanish proficiency and oral fluency were not the only "domain" that had developed in a significant fashion in the immersion context. One of the most important results of the analyses included in Chapter 3 and 4 is the effect that proficiency has played as a common thread in the findings regarding both fluency abilities as well as cognitive abilities.

In Chapter 3 we showed that regardless of level of proficiency IM learners significantly improved across time, particularly in measures of speed of speech delivery. Indeed, this is in live with other findings that showed that low proficiency learners, besides possessing a smaller repertoire of grammatical skills in L2 Spanish, were able to improve their oral skills at a similar rhythm than their high proficiency counterparts. This effect is pervasive in a number of the fluency measures analyzed in this dissertation. For instance, when the amount of speech spoken in the L2 was analyzed, it was seen that IM learners had experienced a significant increase of syllable production across time, however, no effect of proficiency yielded significant at any time point, indicating that both proficiency subgroups had increased their amount of spoken speech at a similar rate. Moreover, when the amount of time spoken in the L2 was also analyzed, it was observed that IM learners had experienced a highly significant increase in their six weeks of immersion. Again, the effect of proficiency was not significant either, which in turn, indicated that both subgroups showed an increase in the amount of time spoken.

On the other hand, a highly significant effect of proficiency was found at Time 1 on rate of speech, one of the global fluency measures analyzed in this dissertation. This significant effect of proficiency was very important, as it indicated the possibility that grammatical knowledge could be related to other aspects of second language development such as oral production (an aspect of SLA that has not been approached in the specialized literature in full extent and that needs further attention). Interestingly, a lack of significant effect of proficiency was observed three weeks after the start of the program (i.e., at Time 2) indicating that in only three weeks both proficiency subgroups were producing a faster rate of speech that was not significantly different from each other. In addition the statistical analysis also indicated that there was a highly significant effect of time on rate of speech for low proficiency learners between Time 1 and Time 2 but not for high proficiency learners, which indicated further that the low proficiency subgroup improved considerably during the first three weeks. Finally, by the end of the program, at Time 3, the effect of proficiency on rate of speech was not very significant ($p=.045$), providing further evidence that there was a significant effect of time for low proficiency learners between Times 2 and 3 (and effect that did not hold for high proficiency learners). This showed that the low proficiency group was able to progressively improve their rate of speech during the six weeks of immersion.

Nonetheless, the analysis of other fluency measures such as the rate of seconds per filled pauses or syllables per filled pause did not reveal a significant effect of proficiency indicating that both subgroups behaved in similar ways with regard to these measures. One difference though is that the IM learners overall did not show any significant effect of time on rate of seconds per filled pause or on the rate of syllables per filled pause. We will not expand on these two measures now as the reason for the lack of significance may lie on the length of the

fragments that were analyzed.

Regarding cognitive abilities, the inclusion of the effect of proficiency also proved to be a fruitful endeavor. For instance, when the Spanish SCT was analyzed, it was observed that there was a highly significant effect of the interaction of time and proficiency which implied that low and high proficiency IM learners developed their L2 Spanish lexical access skills during the six-week period in different ways. In fact, besides their differences in knowledge of Spanish (i.e., in the proficiency test at Time 1), it was observed that in the same way that high proficiency learners became significantly faster accessing new Spanish words, low proficiency learners were also able to significantly become faster across time. Indeed, a lack of effect of proficiency on RT at each data collection time point implied that low proficiency learners were just as cognitively ready at the beginning of the program to benefit from the immersion stay as their high proficiency counterparts. Thus, the individual analysis showed that in the high proficiency learner group (n=13), 6 learners showed significantly faster RT between the beginning and end of IM context and two of them were significantly slower. However, in the low proficiency learner group (n=12), five learners were significantly faster and none of them became significantly slower across time. These results for the Spanish SCT speak for the ability of low proficiency IM learners to improve as much as those that started with higher levels of proficiency at the beginning of the program.

Table 4.11 shows the individual significant degree of accuracy for CONDITION0 and CONDITION1, where 1*** indicates that the individual participant was more accurate in CONDITION1 than CONDITION0, and 0*** implies the opposite. Table 4.11 also presents that individual significant RT results for all cognitive tasks. The task that showed most significant change was the English SCT, as 12 IM learners became slower over time in their L1 during the

immersion experience. As for the English ACT, five learners that were faster in the "non-shift" condition (i.e., CONDITION0) than in the "shift" condition (i.e., CONDITION1) also became faster across time in the Spanish lexical access task, however the sample size is too small to see larger trends between these two cognitive tasks.

Table 4.11. Individual significant RT results for all cognitive tasks

		ENG ACT		SP SCT	ENG SCT	SP PNT
		ACCURACY	RT	RT	RT	RT
PROFICIENCY	PARTICIPANT	COND. 0 vs. 1	COND. 0 vs. 1	T1-T3	T1-T2	T1-T3
LOW	1	**1*****		F***	S*	
LOW	2	0***	F*	F*		
LOW	4	0***				
LOW	8	0***				S**
LOW	9	0***	F*		S***	
LOW	11	**1*****			S**	
LOW	12	0***			S***	F**
LOW	13	**1*****		F***	F***	
LOW	14	0***			S*	
LOW	17	0***			S***	
LOW	21	0***		F***	F***	
LOW	23	0***	F*		F***	
LOW	25	0***	F*		S*	F***
LOW	27	0***		F**	S**	
HIGH	3	1***			F***	
HIGH	5	0***		S**	S***	
HIGH	6	0***			S*	
HIGH	7	0***			F***	S*
HIGH	10	0***		F**		
HIGH	15	0***		S***		
HIGH	16	0***	F*	F*		
HIGH	18	0***				
HIGH	19	0***		F*	S***	
HIGH	20	0***			F***	
HIGH	22	0***	F*	F***		
HIGH	24	0***	F*	F**	F***	
HIGH	26	0***	F*	F***	S*	

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

Another set of critical results derive from the English SCT. As expected, and considering

that this task was prepared with L1 stimuli exclusively, the IM learners did not show any significant effect for proficiency. However, there was a significant effect of the interaction of proficiency and time and indeed, the statistical analysis indicated that while there was no effect of time on RT for high proficiency learners between the beginning and end of the program, there was a significant effect of time on RT for low proficiency learners, who showed an overall lower RT at the end of program than at the beginning. A number of reasons dealing with different hypotheses for language inhibition/attenuation were introduced before; nonetheless, the important point is that the two proficiency groups behaved very differently with the L1 data during the six weeks of immersion. The individual analysis of the English SCT also revealed that eight low proficiency learners became significantly slower and that three low proficiency learners became significantly faster across time. On the other hand, only four high proficiency learners became significantly slower across time while four low proficiency learners became faster between the beginning and end of the program.

As was pointed out in Chapter 1, it is well documented that the L1 as well as the L2 are always activated at the level of the lexicon (Dijkstra, 2005), and even of the grammar (Dussias and Sagarra, 2007). It is also known that bilinguals and second language learners do not use both languages with the same level of frequency and that it is still not perfectly understood how one language is used without the interference of the other. Recent research has shown certain disagreement as to whether the L2 benefits from attenuation or inhibition of the L1 (for a review of inhibitory processes in bilinguals, see Meuter, 2005). Focusing on the IM context at hand, in order to consider that the L1 of a second language learner is inhibited, the inhibited effects should continue once the learner is back in the L2 context (Meuter, 2005). Needless to say, if the measure to establish a real process of L1 inhibition is that the change that learners experienced

during the IM context remain beyond the time of the IM experience, then the design of this dissertation cannot offer a conclusive answer because the data were collected during the abroad experience only.

The results obtained through the English SCT seem to indicate that for low proficiency learners lexical access to English words is more inhibited/attenuated at least through orthography. Quite possibly, the cognitive ability to access L1 words in the mental lexicon through orthographic representation became inhibited/attenuated. However, high proficiency learners, overall, showed a very different outcome in the English and Spanish SCTs, as it seems that they do not change their ability to access L1 words in their mental lexicon. This could be interpreted as an inhibitory/attenuated effect, given that the English SCT was exactly the same at the beginning and at the end of the program. Thus, as was mentioned earlier in this chapter, the fact that high proficiency learners overall do not become statistically faster at Time 2 could be due to some inhibitory/attenuated effect playing equally on their lexical access through visual stimuli (i.e., reading the words included in the task as stimuli).

The results of the Spanish PNT indicated that IM learners kept highly stable results for lexical retrieval, as there was not a significant effect of time on RT either across time or at any data collection time point. Indeed, the lack of effect of proficiency on RT showed that low proficiency learners did not behave differently from high proficiency learners and that their speed and efficiency of lexical retrieval skills for new items did not lag behind those of their higher proficiency counterparts. In sum, the analyses of oral and cognitive data show that although low proficiency learners started with what one could call a certain ‘disadvantage’ with regard to their knowledge of Spanish grammar, they were very able to develop their L2 oral fluency and lexical access and even in lexical retrieval (as they did not show any deleterious

effect or significant difference in comparison to the high proficiency learners at any time point).

CHAPTER 5: CONCLUSION

5.1 Summary of research objectives

This chapter revisits the goals from Chapter 1 and reviews and discusses specific findings as they relate to the overall objectives of this dissertation. The first objective of the dissertation was to quantify the specific differences between the IM and AH learners with regard to the amount of time that learners spent using the L1 and L2 for four communication skills: speaking, writing, reading and listening. The second goal of the dissertation was to investigate the development of second language proficiency and second language fluency in a 6-week IM context and a 15-week AH context. Recall that a unique aspect of the research design was that the same instructor taught the grammar course in both learning contexts and used the same teaching materials. A third goal of this dissertation centered on the investigation of the development of specific cognitive abilities in the IM context. The tasks were two semantic classification tasks (i.e., one in L2 Spanish and one in L1 English) to tap lexical access, a Spanish version of a picture-naming task to tap lexical retrieval, and finally an English attention control task to tap executive attention, the ability to attend to or ignore a given stimulus (i.e., a capacity that has been related to working memory traditionally).

The highlights of this dissertation could be summarized in three main points. First, there is the importance of the longitudinal contribution in which the cognitive data have been presented. This is the first study that approaches oral fluency and cognitive development in an overseas IM learning context to examine data from four different cognitive tasks at three data collection time points for the tasks with Spanish stimuli (i.e., Spanish SCT and PNT) and at two data collection time points for the tasks with the English stimuli (i.e., English SCT and ACT). Second, there is the importance of collecting data *in situ*, an aspect that was also addressed in the

analysis provided in Chapter 2 for the LCP and fluency measures, and which is supported by the results of Chapter 4. Future studies investigating L2 linguistic development specifically and cognitive development at large in different learning contexts should follow this methodological protocol, particularly when collecting sensitive second language data such as the one described in this chapter, which as we have demonstrated can be highly influenced by contextual factors. Third, there is the role of proficiency that was pervasive for fluency and cognitive measures.

5.2 Summary of findings

5.2.1 Language contact profile

Chapter 2 was a detailed analysis of the degree of contact with the target language for the IM and AH learners. This particular set of data originated from two modified versions of the original Language Contact Profile (Freed, Dewey, Segalowitz & Halter, 2004). It is important to recall that some voices have expressed their concern about the amount of hours of L2 contact in SA contexts that second language learners self-report in these questionnaires and that a more 'objective' account should be verified by the host-family members as well as by other people that were in contact with learners while they were abroad such as professionals related to the SA program (Collentine, 2009). Although this is a clear concern for future studies, the principal investigator in this dissertation was a faculty member in both the AH and IM contexts and thus, was a first-hand witness of the amount of exposure to the L2 for learners in each group. Specifically, the IM learners were especially motivated to abide by the language pledge and to communicate exclusively in Spanish (except for the one hour per week in which they were allowed to write to their family members and friends in their native language). At this point it is important to recall the findings of the Language Contact Profile (i.e., LCP) provided in Chapter 2, where it was mentioned that IM learners reported spending most of their time abroad

practicing the four language skills (i.e., speaking, reading, listening and writing) in the target language (i.e., Spanish) and that the use of these skills in their L1 English had been significantly reduced or simply had not occurred. Undoubtedly, the most important aspect of the LCP is the significant contrasts between the number of hours that AH and IM learners reported using Spanish. The disparity between the self-reported amounts of time using the L1 and the L2 by the IM and AH learners is indicative of the fundamental differences between the two learning contexts. In addition to the learning contexts themselves, we may attribute the quantitative differences to the language pledge to which the IM learners abode along with the intensive daily Spanish language curriculum. These two aspects are by far the most distinguishing feature that makes this type of learning context unique within the realm of study abroad contexts.

5.2.2 Proficiency test

As was pointed out in the summary of objectives, the results in this dissertation derive from the comparison of a group of 27 IM learners participating in a six-week IM program and 29 AH learners enrolled in a regular 15-week Advanced Spanish Grammar and Composition in a large Midwestern Institution in the United States. Independently of the differences of both learning contexts and the amount of contact hours with Spanish (whose comparison is one of the motivations for the thorough comparative analysis of both groups carried out in this dissertation), there were no significant differences between the amount of years of instruction previous to learners' participation in the study. This was the reason why the AH group was chosen as the baseline for comparison with the IM group. In addition, after the proficiency test and the oral proficiency task were conducted, it was found that at Time 1, there were not significant effects of GROUP on PROFICIENCY SCORES or on RATE OF SPEECH. These results suggest that both groups were comparable and that the similar amount of prior years of instruction was an

adequate heuristic for assuming comparability between the two groups. The same proficiency test that was administered at the beginning of each context for the IM and AH learners was also administered at the end of the program, and the results indicated that the IM and AH learners had significantly improved their proficiency scores by the end of their respective programs; nonetheless, the degree of significance was higher for the IM learner group than for the AH learner group. Another important aspect of this dissertation derived from the division of each group according to low and high levels of proficiency based on the results obtained at the beginning of the program. Once the IM and AH learner groups were separated according to levels of proficiency, it was seen that while both high and low proficiency IM learners significantly improved their posttest scores, AH learners showed a different outcome. Specifically, the low proficiency AH learners only showed a significant improvement over time whereas high proficiency AH learners did not.

It is challenging to generalize the improvement of the two groups of participants, particularly because the 45-item proficiency test was designed exclusively for this dissertation and has not been used in other studies. However, a number of researchers have studied the development of grammatical abilities in the SA and AH contexts (DeKeyser, 1986, 1990, 1991; Collentine, 2004). DeKeyser found that in both contexts there was almost no difference between the SA and AH contexts, and Collentine (2004) reported that AH learners advance more on aspects such as verbal accuracy and subordinate conjunctions that are addressed in formal instruction, particularly when the subjunctive is introduced. The present investigation did not attempt to replicate the studies of DeKeyser and Collentine, and the instrument to obtain proficiency measures used in this study is rather different. However, it could be said that since the design of this dissertation controlled for using the same grammar materials as well as having

the same grammar instructor in both contexts, the results of the proficiency test in the present investigation indicate that IM learners develop their grammar skills to a higher extent than those in the AH context. A number of studies show findings that learners in the SA context improved their grammatical abilities but these studies did not have an AH control group and focused on specific aspects such as the Spanish copula (i.e., ser/estar) and prepositional collocations (i.e., por/para) (Gutermann, 1992a, 1992b), and improvement of Spanish grammatical abilities through the OPI and SOPI (Isabelle, 2001). Thus, it is problematic to compare our results at this moment; nevertheless, future research needs to address if these specific aspects of grammatical development develop in the same or in a different fashion in intensive IM contexts such as the one described in this dissertation and in SA and AH contexts like the ones investigated in those studies.

5.2.3 Oral second language fluency (speed of delivery and hesitation phenomena)

The third set of results appeared in Chapter 3 and expanded on the performance of second language oral fluency, particularly on two global aspects of oral production, speed of speech delivery (i.e., rate of speech, amount of speech spoken in the L1 and amount of speech spoken in the L2) as well as hesitation phenomena (filled pauses, rate of second per filled pause and rate of syllables per filled pause). All 27 IM learners and 29 AH learners performed two video-retells at the beginning of each learning context (i.e., Time 1), 3 and 7 weeks later respectively (i.e., Time 2), and at the end of each learning context, 6 and 14 weeks later respectively (i.e., Time 3). With regard to speed of speech delivery and as was previously indicated, both groups were highly comparable in terms of rate of speech as the analysis showed that there was not a significant effect of GROUP on RATE OF SPEECH at Time 1. However, across time, the IM learners

showed an increase in rate of speech, whereas the AH learners did not achieve significant changes, which was corroborated by a group effect at Times 2 and 3.

As for measurements of hesitation phenomena, a GROUP effect between IM and AH learner groups was observed for RATE OF SECONDS PER FILLED PAUSE at the beginning of the respective learning experiences. However, a group effect was not found at Times 2 and 3. When the groups were separated according to levels of proficiency, it was seen there was not a significant effect of PROFICIENCY on RATE OF SECONDS PER FILLED PAUSE for either the IM or AH learner groups; however, low proficiency IM learners significantly decreased their rates of seconds per filled pause across time, whereas the high proficiency IM learners did not. Interestingly, AH learners showed a significant effect of PROFICIENCY on the same dependent variable at TIME 1, an effect that disappeared during the following 14 weeks on instruction. Finally, no significant differences were observed for the rate of syllables per filled pause.

These results corroborate findings of previous research on the development of second language fluency in the same context (recall that that the previous research was conducted in years previous to that of the data collection for the current dissertation), specifically, that IM learners increase their rate of speech (García-Amaya, 2008, 2009; García-Amaya & Filgueras-Gómez, 2010). More broadly, they add to the body of research that indicates that overall a SA experience significantly helps to improve rate of speech (Lennon, 1990; Mora & Valls-Ferrer, in press; Segalowitz & Freed, 2004). The different development of fluency experienced by the AH and IM learner groups in their respective contexts is also related to the higher number of hours spent speaking in the L2 (see the results of the LCP in Chapter 2 for a more detailed analysis). Thus, one of the reasons why the amount of L2 practice leads to gains in fluency lies in the procedularization of declarative linguistic knowledge that needs to occur for the delivery of

speech production in an efficient manner. In order to produce output, a number of cognitive processes must evolve from being highly controlled and effortful to automatic and effortless, an aspect that is promoted in the SA context with an increasing number of practice opportunities (DeKeyser, 2001, 2007; Mora and Valls-Ferrer, in press; Segalowitz, 2000, 2004; Segalowitz and Freed, 2004; Vall-Ferrer, 2011). Indeed, L2 models of speech production also offer a plausible explanation for the difference between the AH and IM contexts and the different outcomes that learners participating in each context experience. In her own model, Kormos (2006) attributes lack of fluency to the processing difficulties that L2 learners experience due to the fact that there is a lack of automaticity to encode linguistic knowledge (i.e., lexical, phonological, etc.). In turn, this linguistic knowledge has to be stored and be made available to the language learner as declarative knowledge.

5.2.4 Cognitive abilities (lexical access, lexical retrieval and attention control)

The fourth set of results comes from the analysis of four cognitive tasks that were administered to the IM group only. The IM learner group was the only group that showed a development in fluency measures across time and as such there was an interest to investigate if during the 6 weeks of immersion, these learners had also developed other cognitive abilities such as their ability to access lexical items in English and Spanish in the mental lexicon, to retrieve (access and articulate) Spanish words from their mental lexicon, and to change levels of attention with a task also designed in English. It was hypothesized that the two lexical access tasks and the lexical retrieval task would reflect some level of development (i.e., learners becoming either faster or slower across time); conversely no change was expected in the attention control task. As was expected, the tasks offered insight into the development of second language processing in a six-week IM program from a group as well as an individual perspective. First, IM learners

significantly increased their speed at accessing Spanish words in their mental lexicon over time (i.e., there was a significant effect of TIME between Time 1 and Time 2 and between Time 2 and Time 3), while they also became slower at accessing English words between the beginning and end of the IM context. More importantly, the effect of PROFICIENCY in the English lexical access task unveiled that low proficiency learners only became slower at accessing L1 words in their mental lexicon, an effect that did not appear for the Spanish lexical access tasks where both low and high proficiency learners significantly increased their speed of lexical access. These results for lexical access contrasted with those obtained for lexical retrieval through the picture-naming task. In the latter task, there were no statistically significant differences at any points during the IM program, which indicates that six weeks of immersion were not sufficient to achieve significantly faster lexical retrieval times. This was also confirmed by the individual analysis; recall that between the beginning and end of the program, one participant only became significantly slower and two became significantly faster. This overall result is indicative of two highly differentiated cognitive processes that seem to take place when L2 learners are immersed in a learning context and pledge to use the target language at all times: accessing the mental lexicon (i.e., lexical access) develops in a short period of time whereas accessing and articulating (i.e., lexical retrieval) does not (for new stimuli).

The ACT task indicated that IM learners were not affected by the IM experience in their ability to shift between levels of attention. The degree of accuracy was significantly higher in the "no-shift" condition than in the two "shift" conditions. In addition, the speed of processing indicated that participants processed faster the "no-shift" condition than the "shift" one, an aspect that was also largely supported by the results obtained through the individual analysis. The data for the Spanish lexical access task support Segalowitz and Freed's (2004) findings for faster RTs

for the Spanish stimuli. However, Segalowitz and Freed (2004) also found faster RT for English stimuli, probably due to the fact that data were collected when the learners were back in the United States, and possibly because the learners in their SA program did not have a language pledge. The present findings call into question Segalowitz and Freed's hypothesis of a threshold of cognitive abilities. Indeed, if the threshold exists future research will need to explore how low it needs to be, because it certainly did not apply for the participants of this dissertation. Half of the group that was identified as 'low proficiency' was able to develop their second language skills during the overseas immersion experience, not only in terms of fluency or cognitive abilities, but also in terms of proficiency; if we recall the results of the proficiency test at Time 2, 25 out of 27 learners achieved scores between 25 and 45 points (the range that would grant them high proficiency status), and only two learners scored between 0 and 25 points (the range that would grant them low proficiency status).

In addition, the effects of proficiency on fluency gains in the IM context deserve special attention. Focusing on the IM group and on measurements of speed of speech delivery, there was no effect of PROFICIENCY for any of the following dependent variables related to speed of speech delivery: AMOUNT OF SPEECH SPOKEN IN THE L2 (i.e., syllables) or AMOUNT OF TIME SPOKEN IN THE L2 (i.e., seconds). Considering that there was a significant effect of TIME and that learners spoke more and for longer, this indicates that both groups increased at similar rates. Moreover, and although there was an overall significant effect of PROFICIENCY on RATE OF SPEECH, this effect was barely significant by the end of the program, also indicating that the low proficiency IM learners had consistently improved their rate of speech across time. As for hesitation phenomena, there was no effect of PROFICIENCY on FILLED PAUSE PRODUCTION, RATE OF SECONDS PER FILLED PAUSE or RATE OF

SYLLABLES PER FILLED PAUSE. Certainly, the fact that low proficiency learners improve so much can be considered indirect evidence for the existence of a threshold for learners at lower levels. In any case, the data obtained in this dissertation should be viewed as encouragement to send learners with low levels of proficiency (as low as the ones presented in this dissertation) abroad, because independently of their proficiency level they do respond to the input cognitively. Of course, the data presented in this dissertation show that at the cognitive level low proficiency learners differentiate from high proficiency learners in that they attenuate their L1, which translates in low proficiency learners having a more dramatic change than the high proficiency learners. Recent findings from a group of Catalan-Spanish bilinguals studying L2 English in an SA context in the UK (Valls-Ferrer, 2011) also show that low proficiency learners are those that benefited more from the experience. Thus, it would seem that the threshold of readiness hypothesis (Segalowitz and Freed, 2004; Sunderman and Kroll, 2009) needs to be revisited and probably softened or lowered to accommodate the data by Valls-Ferrer study and the findings of this dissertation.

Another topic explored in the discussion of the cognitive tasks included the possible effects of L1 inhibition/attenuation on the L2. Although the cognitive experiments of this dissertation were not designed to study this particular effect, the English lexical access task showed that once IM learners were separated according to levels of proficiency, those within the low level of proficiency were slower at the end of the IM program (with English stimuli), indicating that such learners were also more affected by L1 inhibition/attenuation. If we consider Levelt's (1989, 1999) model of speech production and attempt to incorporate these findings, two aspects should be pointed out. First, the data from the English lexical access tasks show that grammatical encoding (i.e., the step in which lexical access is embedded), being an earlier

process in the speech chain, is also faster to access by L2 learners and easily inhibited/attenuated in an intensive immersion context. In addition, the experiments indicated that the processes that lead to overt speech after grammatical encoding (i.e., morphophonological encoding, phonetic encoding, and articulation) require almost as much time as lexical access. Indeed, lexical retrieval does not seem to improve easily for L2 learners in barely six weeks of immersion. Whether this attenuation or inhibition is temporary and the extent to which this inhibition takes place for learners with other levels of proficiency are aspects for study in future research.

Among the theoretical hypothesis based on the results obtained in Chapter 4, it is possible to imagine that following Levelt's (1989, 1999) model of speech production for the monolingual mind, for second language learners a six week immersion program would help to improve steps 1 and 2 (i.e., micro and macroplanning and lexical access), however it would not be sufficient for the rest of the components of the model (i.e., morphophonological encoding, phonological encoding and articulation). This finding speaks for the need to provide the field of second language fluency with a theoretical framework to test different hypotheses. Certainly a speech production model for second language learners would be key to carrying out such an endeavor. Up to now, Levelt's model of speech production for monolingual speakers and DeBot's model of speech production for bilingual speakers provide snapshots only of what the ideal monolingual and bilingual speaker would do. On the contrary, the journey of a second language learner towards second language acquisition has multiple caveats and each learner can bring something new into the discussion of how second language acquisition takes place. Thus, a model of speech production needs to account for the variation existing when learners subject themselves to different learning experiences; it should also account for what happens when the learner stops studying the language but still remains a user (i.e., L2 attrition). This hypothetical model would

also need to be flexible and reflect different stages of the acquisitional process.

5.3 Limitations and future directions

In spite of these findings, this investigation has a number of limitations that need to be addressed in future research. Some have been pointed out in the different chapters that comprise this dissertation but here we will summarize those that are most noteworthy. First, it would be important to increase the sample size. Second, it is necessary to add more measures of oral fluency such as length of run between pauses and articulation rate. In addition, it is necessary to include an individual analysis for each of these fluency measures along with measures of accuracy and complexity. Also, the fluency analysis for both AH and IM learners would benefit from other tasks such as a computerized monologic oral production task where participants speak about a variety of topics for a longer period of time. Quite possibly such tasks could provide better insight into the hesitation measures investigated in this dissertation, particularly the rate of seconds per filled pause and the rate of syllables per filled pause. In addition, it would be useful to examine how "stress timed" L1 languages need to become "syllable-timed" L2 with these particular learners, along the lines of Valls-Ferrer (2011) who studies L2 learners of English in an SA context in the UK.

Moreover, there is a need to replicate this study with similar and different populations particularly to investigate if certain effects, such as the one found for the English SCT, are also experienced by other IM learners with similar and/or different levels of proficiency in similar IM contexts as well as in other SA contexts of longer duration. Also, as the data by Segalowitz and Freed show for English (where SA learners became faster with a similar English task), this could be a developmental sequence. Future studies will need to address this issue as well as to analyze cognitive data for the AH learners. Ideally, these future studies will include other cognitive tasks

(such as working memory tasks) and will also examine data in both the L1 and the L2 (for every task including the oral fluency ones). This latter aspect has been highlighted in Segalowitz (2010). Certainly most of these questions cannot be answered by a single experimental design. Still, there is much scope for future research.

5.4 Conclusion

To sum up, this dissertation has offered a thorough description of the use of fluency skills by 27 IM learners who spent six weeks in Spain in summer 2010 in comparison to 29 AH learners who were enrolled in an Advanced Spanish Grammar & Composition course in a large Midwestern institution in Fall 2010. In addition, this dissertation has compared the differences between learner groups for proficiency and fluency development. Moreover, this dissertation has taken into account the development of lexical access, lexical retrieval and attention control. Finally, this dissertation has proven the methodological importance of collecting data *in situ* instead of after learners' return to their country of origin. Overall, the IM learner group has shown a superiority in proficiency and fluency measures over the AH learner group, and this dissertation has tried to explore a number of reasons to explain that situation, but there are others that remain unexplored such as personal motivation. Indeed, the inductive environment toward L2 learning of this particular IM context, the contact hours of formal instruction and the constant exposure to native speech, seem to be the key to improvement for IM learners in general. Unquestionably, future research will need to investigate each of these aspects in isolation in order to understand how much each weighs on the individual learner.

REFERENCES

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica, 128,* 466-478.

Ackerman, P. L. (1988) Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117*, 288–318.

Ackerman, P. L. (1989) Individual differences and skill acquisition. In P. L. Ackerman, R. J. Sternberg, & R. Glaser. (Eds.), *Learning and individual differences: Advances in theory and research*. (pp. 165–217). New York: W. H. Freeman.

Adams. A. M., & Gathercole, S. E. (2000). Limitations in working memory: Implications for language development. *Internalional Journal of Language and Communication Disorders.* 35, 95-116.

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

Atkins, P. W. B., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics, 9*, 537-552.

Averart, S., & Nation, I. S. P. (1991). Fluency improvement in a second language. *The RELC Journal: A Journal of Languge Teaching and Research in Southeast Asia, 22,* 84-94.

Baddeley, A. (1986). *Working memory.* Oxford: Clarendon Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 7-42.

Baddeley. A., Gathercole, S. & Papagno. C. (1998). 'The phonological loop as a language learning device. *Psychological Review, 105,* 158-173.

Binnenporte, D., Van Bael, C., den Os, E., & Boves, L. (2005). Gender in everyday speech and language: A corpus-based study. *Interspeech,* 1-4.

Blake, J., Austin, W., Cannon, M., Usus, A., & Vaughan, A. (1994). The relationship between memory span and measures of imitative and spontaneous language complexity in preschool children. *International Journal of Behavioral Development.n, 91-107.*

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: effects of relationship, topic, role, and gender. *Language and Speech, 44*, 123-147.

Brecht, R., Davidson, D., Ginsberg, R. (1990). The empirical study of proficiency gain in study abroad environments of American students of Russian. In *American contributions to the VII Internation Congress of MAPRIAL* ed. By D. Davidson, 123-152. Washington, D.C.: American Council of Teachers of Russian.

Brecht, R., Davidson, D., Ginsberg, R. (1993). *Predictors of foreign language gain during study abroad.* Washington D. C.: National Foreign Language Center.

Caramazza, A., & Brones, I. (1980). Semantic classification by bilinguals. *Canadian Journal of Psychology. 34*, 1, 77-81

Carroll, J. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals, 1,* 131-151.

Czerwionka, Lori & Lorenzo García-Amaya. (2011). Variability and clausal position of discourse markers in L2 oral production: Sociopragmatic learning in a 6-week overseas immersion program. *Hispanic Linguistics Symposium.* Athens, GA, October 6-9, 2011.

Coleman, J. (1995). A comparative survey of the proficiency and progress of language learners in British universities. *Der C-Test. Theoretische Grundlagen und prakische Anwendungen Volume 3,* ed. By R. Grotjahn. Bochum: Brockmeyer.

Chambers, F. (1997). What do we mean by fluency? *System, 25*, 535-544.

Cheung. H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology, 32*, 867-873.

Clark, H., & Fox Tree, J., (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73-111.

Cohen, J., Aston-Jones, G., & Gilzenrat, M. S. (2004). A systems-level perspective on attention and cognitive control. In M. Poner (Ed.), *Cognitive neuroscience of attention* (pp. 71-90). New York: The Guilford Press.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257–271.

Coleman, J. A. (1997). Residence abroad within language study. *Language Teaching*, 30, 1-20.

Coleman, J. A. (1998). Language Learning and Study Abroad: the European Perspective. *Frontiers, 4*, 167-203.

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition, 26*, 227-248.

Collentine, J. (2009). Study abroad research: Findings, implications and future directions. In C. Doughty, & M. Long (Eds.), *Handbook of Language Teaching* (pp. 218-233). Malden, MA: Blackwell.

Collentine, J., & Freed, B. (2004). Introduction: Learning context and its effects on second language acquisition. *Studies in Second Language Acquisition, 26*, 153-172.

Costa, A. (2005). Lexical access in bilingual production. In J. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic perspectives* (pp. 308-325). New York: Oxford University Press.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language, 41,* 365-397.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language, 50,* 491-511.

Costa, A., & Santesteban, M. (2006) (Guest Eds.). Lexical access in bilingual speech production. Special issue of *Bilingualism: Language and Cognition, 9.*

Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America, 107,* 989-999.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America*, *111*, 2862–2873.

Czerwionka, L. & García-Amaya, L. (2011). Variability of discourse markers in L2 oral production: Sociopragmatic learning in a 6-week overseas immersion program." *Hispanic Linguistics Symposium*, Athens, GA. October 6-9, 2011.

Darcy, I., Park, H-P., Yang, C.-L., and Gleiser, A. (2012) Individual differences in the development of L2 phonological processing: The contribution of cognitive abilities and executive function. Poster presented at the 1st Conference on Sources of Individual Linguistic Differences. University of Ottawa, Canada, March 2-4, 2012.

De Bot, K. (1992). A bilingual production model: Levelt's 'speaking' mode adapted. *Applied Linguistics, 13*, 1-24.

De Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning, 46*(3), 529-555.

De Bot, K. (2003). Bilingual speech: from concepts to articulation. *Fremdsprachen Lehren und*

Lernen, 32, 91-103.

De Bot, K. (2004). The multilingual lexicon: modeling selection and control. *International Journal of Multilingualism, 1,* 17-32.

De Bot, K. & Schreuder, R. (1993). Word production and the bilingual lexicon. In R.Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon* (pp. 191-214). Amsterdam: Benjamins.

De Jong, N. H. Wempe, T. (2007). Automatic measurement of speech rate in spoken Dutch. *ACLC Working papers, 2*(2), 49-58.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei ad measure speech rate automatically. *Behavior Research Methods, 41,* 385-390.

Dechert, H. W. (1980). Pause and intonation as indicators of verbal planning in second-language speech productions: Two examples from a case study. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 271-285). The Hague: Mouton.

Dechert, H. W., Möhle, D., Raupach, M. (Eds.) (1984). *Second language productions.* Tübingen, Germany; Gunter Narr Verlag.

Dechert, H. W., & Raupach, M. (Ed.). (1980). *Temporal Variables in Speech*. Mouton: The Hague.

Dechert, H. W. & Raupach, M. (Eds.) (1987). *Psycholinguistic models of production.* Norwood, NJ; Ablex Publishing Corporation.

DeKeyser, R. (1986). From learning to acquisition? Foreign language development in a US classroom and during a semester abroad. Stanford, CA: Stanford University dissertation.

DeKeyser, R. (1990). From learning to acquisition? Monitoring in the classroom and abroad. *Hispania, 73*, 238-47.

DeKeyser, R. (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. (1997). Beyond explicit rule learning: Automatizing second language morpho-syntax. *Studies in Second Language Acquisition, 19*, 195-221.

DeKeyser, R. (2001) Automaticity and automatization. In P. Robinson. (Ed.), *Cognition and second language instruction.* (pp. 125–151). New York: Cambridg University Press.

DeKeyser, R. (2010). Monitoring processes in Spanish as a second language during a study abroad program. *Foreign Language Annals, 43*, 80-92.

Derwing, T., & Munro, M.(1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*, 1-16.

Derwing, T. M., & Munro, M.J. (2001). What speaking rates do nonnative listeners prefer? *Applied Linguistics, 22*, 324–37.

Derwin, T., & Rossiter, M., Munro, M., & Thomson, R. (2004). Second language fluency: Judgments on different tasks. *Language Learning, 54,* 655-679.

Derwin, T., Thomson, R., & Munro, M. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System, 34,* 183-193.

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition, 26*, 249-273.

Díaz-Campos, M. (2006). The effect of style in second language phonology: an analysis of segmental acquisition in study abroad and regular-classroom students.

Dörnyei, Z., Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty& M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Malden, MA: Blackwell.

Ejzenberg, R. (1992). *Understanding nonnative oral fluency: The role of task structure and discourse variability*. Unpublished doctoral dissertation. State University of New York, Albany.

Ellis, R. (2003). Tasks, production, and language acquisition. In R. Ellis (Ed.), *Task-based language learning and teaching* (pp. 103-139): Oxford University Press.

Ellis, R., & Yuan, F. (2005). The effects of careful within-task planning on oral and written task performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 167-192). Amsterdam: John Benjamins.

Fillmore, C. (1979). On fluency. In C. J. Fillmore, D. Kempler, & W. S. Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85–101). New York: Academic Press.

Firth, A., & Wagner, J. (1997). On discourse, communication and (some) fundamental concepts in SLA research. *Modern Language Journal 81*, 285-300.

Favreau, M., & Segalowitz, N. (1983). Automatic and controlled processes in the first and second language reading of fluent bilinguals. *Memory & Cognition, 11,* 565-674.

Flege, J. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language*

research (pp. 229-273). Timonium, MD: York Press.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21,* 354-375.

Forster, K. I., & Forster, J. C. (2003). DmDx: A window display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers, 35*, 116-124.

Foster, P. & Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second language Acquisition*, 18, 299-324.

Fox Tree, J. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition, 29.* 320-326.

Fox Tree, J. (2002). Interpreting pauses and ums at turn exchanges. *Discourse Processes, 34,* 37-55.

Freed, B. (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. Atlatis (Ed.), *Linguistics, language teaching and language acquisition: the interdependence of theory, practice and research (GURT 1990)*. Washington, D.C.: Georgetown University Press.

Freed, B. F. (Ed.). (1995a). *Second Language Acquisition in a Study Abroad Context*. Amsterdam and Philadelphia: John Benjamins.

Freed, B. F. (1995b). 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 and Philadelphia: John Benjamins.

Freed, B. (1998). An overview of issues and research in language learning is a study abroad setting. *Frontiers, 4*, 31-60.

Freed, B. (2000). Is fluency, like beauty, in the eyes (and ears) on the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243-265). Ann Arbor, University of Michigan Press.

Freed, B. (2008). Second language learning in a study abroad context. In N. Van Deusen-Scholl and N. H. Hornberger (eds), *Encyclopedia of Language and Education*, 2nd Edition, Volume 4: Second and Foreign Language Education, 113–125.

Freed, B., Dewey, D., Segalowitz, N., Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition, 26*, 349-356.

Freed, B., Segalowitz, N., Dewey, D. (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.

Freed, B. F., So, S., & Lazar, N.A. (2003). Language learning abroad: how do gains in written fluency compare with oral fluency in French as a second language? *ADFL Bulletin, 43*, 34-40.

Freed, B. (2008). Second language learning in a study abroad context. In N. Van Deusen-Scholl and N. H. Hornberger (eds), *Encyclopedia of Language and Education*, 2nd Edition, Volume 4: Second and Foreign Language Education, 113–125.

García-Amaya, L. (2006). Disfluencies in advanced second language learners: A drawback or a strategy? *Hispanic Linguistic Symposium*, London, Ontario. October 19-22, 2006.

García-Amaya, L. (2008). The effect of topic on rate of speech. *ELIA, 8*, 117-150.

García-Amaya, L. (2009). New findings on fluency measures across three different learning contexts. In Collentine, J., García, M., Lafford, B., & Marcos-Marín, F. (Eds.) *Selected Proceedings of the 11th Hispanic Linguistics Symposium*, (pp. 68-80). Somerville, MA: Cascadilla Press.

García-Amaya, L. (2010). *The effect of clause type on filled pauses and discourse markers clause type in the oral speech production of second language learners across three different learning contexts*. Kentucky Foreign Language Conference. Lexington, KY, April 15- 17, 2010.

García-Amaya, Lorenzo & Marisa Filgueras. (2010). From filled pauses to discourse markers in 6 weeks in an overseas immersion program: An account of the relationship of these particles with clause type and oral fluency. *Hispanic Linguistics Symposium*, Bloomington, IN, October 14-17, 2010.

García-Amaya, Lorenzo. (2011a). The effect of clause type on filled pauses and discourse markers in the oral production of second language learners across three different learning contexts. *American Association for Applied Linguistics.* Chicago, IL, March 26-29, 2011.

García-Amaya, Lorenzo. (2011b). Cognitive demands of the study abroad experience. *Second Language Research Forum.* Ames, IA, October 13-16, 2011.

Gass, S. (1997). *Input, interaction, and the second language learner.* Mahwah, NJ: Lawrence Erlbaum Associates.

Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly, 22,* 473-492.

Gatbonton, E. & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review, 61,* 325-353.

Gatbonton, E., & Trofimovich, P. (2008). The ethnic group affiliation and L2 proficiency

link: Empirical evidence. *Language Awareness, 17,* 229-248.

Ginsberg, R. (1992). Language gains during study abroad: An analysis of the ACTR data. National Foreign Language Center Working Papers. Washington D. C.: National Foreign Language Center.

Goldman-Eisler, F. (1951). The measurement of time sequences in conversational behavior. *British Journal of Psychology, 42,* 355-362.

Goldman-Eisler, F. (1961). Hesitation and information in speech. In C. Cherry (Ed.), *Information theory* (pp. 162-174). London: Butterworths.

Goldman-Eisler, F. (1968). *Psycholinguistics experiments in spontaneous speech.* London: Academic Press.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and speech, 15,* 103-113.

Green, D. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1,* 67-81.

Griffiths, R. (1991). Pausological research in an L2 context: A rationale, and review of selected studies, *Applied Linguistics, 12,* 345-364.

Grosjean, F. (1980a). Linguistic structures and performance structures: Studies in pause distribution. In H. W. Dechert & M. Raupach, (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 91-106). The Hague: Mouton.

Grosjean, F. (1980b). Temporal variables within and between languages. In H. W. Dechert & M. Raupach, (Eds.), *Towards a cross-linguistic assessment of speech production* (pp. 39-53). Frankfurt: Lang.

Guntermann, G. (1992a). An analysis of interlanguage development over time: part I, *por* and *para*. *Hispania 75,* 177-87.

Guntermann, G. (1992b). An analysis of interlanguage development over time: part II, *ser* and *estar*. *Hispania 75*, 1294-1303.

Guntermann, G. (1995). The Peace Corps experience: language learning in training and in the field. In Freed, B. F. *Second language acquisition in a study abroad context*. (pp. 149-69). Philadelphia: John Benjamins.

Hashimoto, H. (1994). Language acquisition of an exchange student within the homestay environment. *Journal of Asian Pacific Communication, 4,* 209-224.

Huebner, T. (1991). Effects of overseas study: A preliminary report on an intensive beginning-level course in Japanese. Paper presented at the annual PCIE Conference, Pittsburgh, PA.

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.

Isabelli, C. (2001). Motivation and extended interaction in the study abroad context: factors in the development of Spanish language accuracy and communication skills. Austin: University of Texas dissertation.

Isabelli, C. (2004). The acquisition of the null subject parameter properties in SLA: Some effects of positive evidence in a naturalistic learning context. *Hispania 87*, 150-162.

Flege, J. E., Munro, M. J., & Mackay, I. R. A. (1995) Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*, 3125–3134.

Koponen, M., & Riggenbach, H. (2000) Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), *Perspectives on fluency*. (pp. 5–24) Ann Arbor: University of Michigan Press.

Kormos, J. (1999). Monitoring and self-repair in L2. *Language Learning, 49*, 303-42.

Kormos, J. (2006). *Speech production and second language acquisition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Kroll, J., & de Groot, A. M. B. (Eds.) (2005). *Handbook of bilingualism: Psycholinguistic approaches.* Oxford, UK: Oxford University Press.

Kroll, J., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research, 18*, 137–171.

Kroll, J. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connection between bilingual memory representations. *Journal of Memory and Language, 33,* 149-174.

Kroll, J. & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In J. Kroll & A. M. B. de Groot (Eds.). *Handbook of bilingualism: Psycholinguistic perspectives* (pp. 531-553). New York: Oxford University Press.

Lafford, B. A. (2004). The effect of context of learning on the use of communication strategies by learners of Spanish as a second language. *Studies in Second Language Acquisition 26*. 201-26.

Lafford, B. A. (2006). The Effects of Study Abroad vs. Classroom Contexts on Spanish SLA: Old Assumptions, New Insights and Future Research Directions. In Klee, C.A., & Face, T. L. (Eds.). *Selected Proceedings of the 7th Conference on the Acquisition of Spanish*

and Portuguese as First and Second Languages (pp. 1-25). Somerville, MA: CascadillaProceedings Project.

Lafford, B. A., Collentine, J. G. (2006). The effects of study abroad and classroom contexts. In R. Salaberry, Lafford, B. A. (Ed.), *The Art of Teaching Spanish* (pp. 103-126). Washington, D.C.: Georgetown University Press.

Llanes, A., Muñoz, C. A Short stay abroad: Does it make a difference? *System, 37,* 353-365.

Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics.* Oxford, UK: Oxford University Press.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40,* 387-417.

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25-42). Ann Arbor: The University of Michigan Press.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown ad P. Hagoort (Eds.), *The neurocognition of language* (pp. 83-22). Oxford, UK: Oxford University Press.

Levelt, W., Roelofs, A. & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1-75.

Lord, G. (2006). Defining the indefinable: study abroad and phonological memory abilities. In Klee, C.A., & Face, T. L. (Eds.). *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 40-46). Somerville, MA: CascadillaProceedings Project.

Lord, G. (2009). Secong language awareness and developmental during study abroad: A case study. *Hispania, 92,* 136-150.

Llanes, À. (2011). The many faces of study abroad: an update on the research on L2 gains emerged during a study abroad experience. *International Journal of Multilingualism*.

Llanes, À. & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, 37 (3), 353-365.

MacWhinney, B. (2008). A unified model. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 341-371). New York: Routledge.

Masoura, E. V. & Gathercole, S. E. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory, 3,* 422~29.

Meara, P. (1993). The bilingual lexicon ad the teaching of vocabulary. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 279-297). Amsterdam: John Benjamins.

Meara, P. (1994). The year abroad and its effects. *Language Learning Journal, 10,* 32-38.

Meara, P. (1994) Second language acquisition: Lexis. In Asher, R.E. (Ed.) *The encyclopedia of language and linguistics. Oxford*: Pergamon. Vol. 7. 3276-3228.

Meuter, R. (2005). Language selection in bilinguals: mechanisms and processes. In J. Kroll & A. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 349-370). Oxford, Uk: Oxford University Press.

Moder, C. L., Martinovic-Zic, A. (2004). *Discourse Across Languages and Cultures*. John Benjamins Publishing Company

Möhle, D. (1984). A comparison of the second language speech of different native speakers. In H. W. Dechert, Möhle, D., Raupach, M. (Ed.), *Second language productions* (pp. 26-49). Tubingen: Gunter Narr.

Mora, J. C., & Valls-Ferrer, M. (in press). Oral Fluency, Accuracy and Complexity in Formal Instruction and Study Abroad Learning Contexts. *TESOL Quarterly*.

Munro, M. J., & Derwing, T. M., (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing, 3*, 253-266.

Munro, M. J., & Derwing, T. M., (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*, 289-306.

Munro, M. J., & Derwing, T. M., (1999) Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 49*, 285-310.

Munro, M. J. , & Derwing, T. M., (2001) Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, *23*, 451–468

Nation, I.S.P. (1989). Improving speaking fluency. *System, 17,* 377-384.

O'Brien, I., Segalowitz, N., Collentine, J., Freed, B. F. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics, 27,* 377-402.

O'Brien, I., Segalowitz, N., Freed, B. F., Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition, 29*, 557-581.

Olynyk, M., d'Anglejan, A., & Sankoff, D. (1990). A quantitative and qualitative analysis of speech markers in the native and second language speech of bilinguals. In R. Scarcella, R. Andersen & S. Krashen (Eds.), *Developing communicative competence in a second language* (pp. 139-155). Rowley, MA: Newburry House.

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition, 21,* 109-148.

Ortega, L. (2005). What do learners plan: Learner-driven attention to form during pre-task planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 77-109). Amsterdam: John Benjamins.

Ortega, L. (2007). Meaningful L2 practice in foreign language classrooms: A cognitive-interactionist SLA perspective. In R. DeKeyser (Ed.), *Practice in a second language* (pp. 180-207). Cambridge, UK: Cambridge University Press.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism.* Amsterdam: John Benjamins.

Perfetti, C. A. (1985) *Reading ability*. Oxford: Oxford University Press.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics, 15*, 36–57.

Raupach, M. (1980). Temporal variables in first and second language speech production. In H. W. Dechert, Raupach, M. (Ed.), *Temporal variables in speech* (pp. 49-60). The Hague: Mouton.

Raupach, M. (1984). Formulae in second language speech production. In H. W. Dechert, D. Mohle & M. Raupach (Eds.), *Second language productions* (pp, 114-137). Tübingen: Gunter Narr.

Raupach, M. (1987). Procedural knowledge in advanced learners of a foreign language. In J. Coleman, & R. Towell (Eds.), *The Advanced Language Learner* (pp. 123-157). London: AFSL/CILT.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of non-native speaker conversations. *Discourse Process, 14,* 423-441.

Riggenbach, H. (Ed.) (2000). *Perspectives on fluency.* Ann Arbor: University of Michigan Press.

Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning*, 45, 99-140.

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. (Ed.) (2002). *Individual differences and instructed language learning.* Amsterdam: John Benjamins.

Ryan, J. & Lafford, B. (1992). Acquisition of lexical meaning in a study abroad environment: ser and estar and the Grenada experience. *Hispania, 75,* 714-722.

Sajavaara, K. (1987). Second language speech production: Factors affecting fluency. In J. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 45-65). Norwood, NJ: Ablex Publishing Corporation.

Sanfeliu, M. C., & Fernandez, A. (1996). A set of 254 Snodgrass–Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, and Computers, 28*, 537–555.

Segalowitz, N. (1997). Individual differences in second language acquisition. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 85–112). Mahwah, NJ: Erlbaum.

Segalowtiz, N., Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish at home and study abroad *Studies in Second Language Acquisition, 26*, 173-199.

Segalowitz, N., Freed, B. F., Collentine, J. G., Lafford, B., A., Lazar, N., Díaz-Campos, M. A. (2004). A comparison of Spanish second language acquisition in two different learning contexts: study abroad and the domestic classroom. *Frontiers, 10*, 21-38.

Segalowitz, N., O'Brien, I., & Poulsen, C. (1998). Evidence for a domain-specific component of attentional control in skilled performance. *Brain and Cognition, 37*, 129-132.

Segalowitz, N., & Segalowitz, S. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*, 369–385.

Segalowitz, N., Trofimovich, P., Gatbonton, E., Sokovskaya, A. (2008). Feeling affect in a second language: The role of word recognition automaticity. *The Mental Lexicon, 3,* 47-71.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology, 45,* 21-50.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49,* 93-120.

Schiffrin, D. (1988). *Discourse Markers*. Cambridge University Press

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14,* 357-385.

Schmidt, R. (1995). *Attention and awareness in foreign language learning.* Honolulu: University of Hawaii Press.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge, UK: Cambridge University Press.

Shebilske, W., Goettl, B., & Regian, J. W. (1999). Executive control of automatic processes as complex skills develop in laboratory and applied settings. In D. Gopher & A. Koriat. (Eds.), *Attention and performance XVII: Cognitive regulation of performance—Interaction of theory and application* (pp. 401–432). Cambridge, MA: MIT Press.

Snodgrass, J. G., &Vanderwart, M. (1980). Standardized set of 260 picture-norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174–215.

Sunderman, G., & Kroll, J. F., When study-abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics, 30,* 79-99.

Simões, A. (1996). Phonetics in second language acquisition: An acoustic study of fluency in adult learners of Spanish. *Hispania, 79*, 87-95.

Stroop, J. (1935). Studies of interference in serial and verbal reactions. *Journal of Experimental Psychology, 18,* 643-662.

Sunderman, G. (2002). A psycholinguistic investigation of second language lexical processing. Unpublished doctoral dissertation, Pennsylvania State University.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471-483). Mahwah, NJ: Lawrence Erlbaum Associates.

Talburt, S., & Stewart, M. A. (1999). What's the subject of study abroad?: race, gender, and "living culture." *Modern Language Journal 83*,163-75.

Taube-Schiffnorman, M., Segalowitz, N., (2005). Within-language attention control in second language processing. *Bilingualism: Language and Cognition, 8*,195-206

Temple, L. (1985). He who hesitates is not lost: Fluency and the language learner. *Revue de*

Phonetique Appliquee, 73-74-75, 293-302.

Temple, L. (1992). Disfluencies in learner speech. *Australian Review of Applied Linguistics, 15*, 29-44.

Temple, L. (2000). Second language learner speech production. *Studia Linguistica, 54,* 288-297.

Temple, L. (2005). Instructed learners' fluency and implicit/explicit language processes. In A. Housen & M. Pierrard (Eds.) *Investigations in Instructed Second Language Acquisition.* (pp. 31-49). Berlin: Mouton de Gruyter.

Tokowicz, N., Michael, E. B., & Kroll, J. F. (2004). The roles of study-abroad experience and working memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition, 7*, 255–272.

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

Towell, R. (1987). Approaches to the analysis of the oral language development of the advanced learner. In J. Coleman, A., Towell, R. (Ed.), *The advanced language learner: papers of the Joint AFLS/SUFLRA Conference* (pp. 157-182). London: Association for French Language Studies in association with CILT.

Towell, R. Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17,* 84-119.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology, 49*, 51–79.

Wiese, R. (1984). Language production in foreign and native languages: same or different? In H. W. Dechert, Möhle, D., Raupach, M. (Ed.), *Second language productions* (pp. 11-17). Tubingen: Gunter Narr.

Wood, D. (2001). In search of fluency: What is it and how can we teach it? *Canadian Modern Language Review, 57*, 573–589.

Wood, D. (2006). Uses and Functions of Formulaic Sequences in Second Language Speech: An Exploration of the Foundations of Fluency. *The Canadian Modern Language Review, 63*, 13-33.

Woolsey, D. (2006). Second language acquisition of the verb *estar* with adjectives: An exploration of contexts of comparison and immediate experience. Unpublished doctoral dissertation, Indiana University, Bloomington.

Yuan, F., Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics, 24*, 1-27.

APPENDIX A: PRETEST (PARTS 1 & 2) AND POSTTEST VERSIONS OF THE LANGUAGE CONTACT PROFILE

FIRST VERSION OF THE LANGUAGE CONTACT PROFILE

PART 1: Background Information

1. Gender: Male / Female

2. Age:_

3. Country of birth: _____ _

4. What is your native language? 1) English 2) Spanish 3) Other _____ _

5a. What language(s) do you speak at home? 1) English 2) Spanish 3) Other _____ _

5b. If more than one, with whom do you speak each of these languages? _____ _

6a. ln what language(s) did you receive the majority of your *high school* education?

1) English 2) Other _____ _

6b. If more than one please give the approximate number of years for each language. ____

7a. Have you ever visited a Spanish-speaking region *for the purpose of studying Spanish* before this semester*?*
Circle one: Yes / No

7b. If yes, when? Where? _____ ___

7c. For how long? __1 semester or less -____2 semesters –or more ___than 2 semesters

8. Other than the experience mentioned In Question 7, have you ever lived in a situation where you were exposed to a language other than your native language (e.g., by living in a multilingual community; visiting a community for purposes of study abroad or work; exposure through family members, etc.)? Circle one: Yes / No

If Yes, please give details below. If more than three, list others on back of this page.

	Experience 1	Experience 2	Experience 3
Country/region			
Language			
Purpose			
From when to when			

9. In the boxes be1ow, rate your language ability in each of the languages that you know. Use the following ratings: 0) Poor, 1) Good, 2) Very good, 3) Native/nativelike.
How many years (if any) have you *studied* this language in a *formal school* setting?

Language	Listening	Speaking	Reading	Writing	Number of years of study
English					
Spanish					
Other					

10. Have you studied Spanish in school in the past at each of the levels listed below? If yes, for how long?

a) Elementary school: _No _Yes: _less than 1 year _1-2 years _more than 2 years

b) Junior high (mid.) school: _No _Yes: _less than 1 year _1-2 years _more than 2 years

c) Senior high school: _No _Yes: _less than 1 year _1-2 years _more than 2 years

d) University/college: _No _Yes: _less than 1 year _1-2 years _more than 2 years

e) Other (Please specify) _____

_No _Yes: _less than 1 year _1-2 years _more than 2 years

11. What year college did you finish in June 2010l? (circle one):

Freshman

Sophomore

Junior

Senior

Other

PART 2: All of the Questions That Follow Refer to Your Use of Spanish, Not Your Native Language (Unless the Question Says Otherwise)

13. On average, how often did you *communicate* with native or fluent speakers of Spanish *in Spanish* in the year prior to the start of the current semester?

0) never 1) a few times a year 2) monthly 3) weekly 4) daily

14. Use this scale provided to rate the following statements.

0) never 1) a few times a year 2) monthly 3) weekly 4) daily

Prior to this current semester, I tried to speak Spanish to:

_ a. my instructor outside of class

_ b. friends who are native or fluent speakers of Spanish

_ c. classmates

_ d. strangers whom I thought could speak Spanish

_ e. a host family, if living in a Spanish-speaking area

_ f. service personnel (e.g., bank clerk, cashier)

15. For each of the items below, choose the response that corresponds to the amount of time you estimate that you spent on average doing each activity *in Spanish* prior to the start of the current semester.

a. watching Spanish language television

0) never 1) a few times a year 2) monthly 3) weekly 4) daily

b. reading Spanish language newspapers

0) never 1) a few times a year 2) monthly 3) weekly 4) dally

c. reading novels in Spanish

0) never 1) a few times a year 2) monthly 3) weekly 4) daily

d. listening to songs in Spanish

0) never 1) a few times a year 2) monthly 3) weekly 4) daily

e. reading Spanish language magazines

0) never 1) a few times a year 2) monthly 3) weekly 4) daily

f. watching movies or videos in Spanish

0) never 1) a few times a year 2) monthly 3) weekly 4) daily

16. List any other activities that you commonly did using Spanish prior to the start of the current semester

17. Please list all the Spanish courses you were taking during the last academic year. This includes Spanish language courses as well as content area courses taught in the Spanish language.

Course name	Brief description
_____	_____
_____	_____
_____	_____
_____	_____

POSTTEST VERSION OF THE LANGUAGE CONTACT PROFILE

Name: _____

Please indicate the Spanish language courses you are taking at NIU:

Course name and brief description

_____ _____

_____ _____

_____ _____

_____ _____

1. Which situation best describes your living arrangement during the semester?

a. _ (I lived in the home of a Spanish-speaking family)

i. List the members of the family (e.g., mother, father, one 4-year-old daughter)

_____

ii. Did they speak English frequently? Circle one: Yes / No

iii. Were there other nonnative speakers of Spanish living with your host family? Circle one: Yes / No

b. _ I lived in the student dormitory.

i. _ I had a private room.

ii. _ I had a roommate who was a native or fluent Spanish speaker.

iii. _I lived with others who are NOT native or fluent Spanish speakers.

c. _ I lived alone in a room or an apartment.

d. _ I lived in a room or an apartment with native or fluent Spanish speaker(s).

e. _ I lived in a room or an apartment with others who are NOT native or fluent Spanish speakers.

f. _ Other. Please specify: _____ _

For the following items, please specify:

(i) How many *days per week* you typically used Spanish in the situation indicated, and

(ii) on average how many *hours per day* you did so.

Circle the appropriate numbers.

2. On average, how much time did you spend speaking, *in Spanish,* outside of class with native or fluent Spanish speakers during the semester?

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

3. During the semester, outside of class, I tried to speak *Spanish to:*

3a. my instructor

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

3b. friends who are native or fluent Spanish speakers

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

3c. classmates

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

3d. strangers whom I thought could speak Spanish

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

3e. a host family, Spanish roommate, or other Spanish speakers in the dormitory

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

3f. service personnel,

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

3g. other; specify:

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

4. How often did you use Spanish outside the classroom for each of the following purposes?

4a. to clarify classroom-related work

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

4b. to obtain directions or information (e.g., "Where is the post office?", "What time is the train/bus to …?", "How much are stamps?")

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

4c. for superficial or brief exchanges (e.g., greetings, "Please pass the salt," "I'm leaving," ordering in a restaurant) with my host family, roommate, or acquaintances in a dormitory

TypicaI1y, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

4d. extended conversations with my host family, roommate, friends, or acquaintances in a dormitory, native speakers of English with whom I speak Spanish.

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

5a. How often did you try deliberately to use things you were taught in the classroom (grammar, vocabulary, expressions) with native or fluent speakers outside the classroom?

Typically. how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

5b. How often did you take things you learned outside of the classroom (grammar. vocabulary, expressions) back to class for question or discussion?

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

6. How much time did you spend doing the following *each week?*

6a. speaking a language other than English or Spanish to speakers of that language (e.g., Chinese with a Chinese-speaking friend)

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

6b. speaking *Spanish* to native or fluent speakers of *Spanish*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

6c. speaking *English* to native or fluent speakers of *Spanish*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

6d. speaking *Spanish* to nonnative speakers of *Spanish* (i.e., classmates)

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

6e. speaking *English* to nonnative speakers of *Spanish* (I.e., classmates)

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

7. How much time did you spend doing each of the following activities *outside of class?*

7a. overall, in reading in Spanish *outside of class*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

7b. reading Spanish newspapers *outside of class*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

7c. reading novels in Spanish *outside of class*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

7d. reading Spanish language magazines *outside of class*

Typically, how many *days per week?* 0 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

7e. reading schedules, announcements, menus, and the like in Spanish *outside of class*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

7f. reading e-mail or Internet web pages in Spanish *outside of class*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

7g. overall, in listening to Spanish *outside of class*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

7h. listening to Spanish television and radio *outside of class*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

7i. listening to Spanish movies or videos *outside of class*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

7j.listening to Spanish songs *outside of class*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

7k. trying to catch other people's conversations in Spanish *outside of class*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

71. overall, in writing in Spanish *outside of class*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

7m. writing homework assignments in Spanish *outside of class*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

7n. writing personal notes or letters in Spanish *outside of class*

Typically, how many *days per week?* 0 l' 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

70. writing mail in Spanish *outside of class*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

7p. filling in forms or questionnaires in Spanish *outside of class*

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

8. On average, how much time did you spend speaking in *English* outside of class during the semester?

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

9. How often did you do the following activities in *English* during the semester?

9a. reading newspapers, magazines, or novels or watching movies, television, or videos

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

9b. reading e-mail or Internet web pages in English

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

9c. writing e-mail in English

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

9d. writing personal notes and letters in English

Typically, how many *days per week?* 0 1 2 3 4 5 6 7

On those days, typically how many *hours per day?* 0-1 1-2 2-3 3-4 4-5 more than 5

THAT'S ALL! THANK YOU!

APPENDIX B: APPLICATION PROCESS TO THE INDIANA UNIVERSITY HONORS PROGRAM IN FOREIGN LANGUAGES FOR HIGH SCHOOL STUDENTS[1]

Application Process

The IUHPFL attracts a large number of candidates and is very competitive. For that reason, the application process is very thorough and involves several steps:

Step One: Testing

- Eligible students interested in applying for the IUHPFL will first need to take the language test, given on regional sites in the state of Indiana in mid-October.
- Students wishing to take the test will need to fill out the test registration information and make the test fee payment via credit or debit card. Only students who have submitted the test registration form and paid the test fee will be allowed to take the test.
- The language test includes assessment in grammar, reading and listening skills. During the testing, students will also submit an authentic writing sample on a given topic.

Step Two: Application

- Students who have scored high enough on the language test are invited to apply to the IUHPFL. Before making application, we strongly encourage you to take the IUHPFL Self-Inventory to see if the Honors Program is the right study abroad program for you.
- The IUHPFL selection committee uses a detailed rubric to evaluate each candidate for the Honors Program. The rubric includes sub-sections dedicated to assess your language test scores, official school transcripts, your personal application as well as confidential appraisals by teachers and non-related adults.

[1] Taken verbatim from http://www.indiana.edu/~iuhpfl/applicationprocess/

- Spanish students will have a chance on the Personal Application to list a preference to study in Mexico or in Spain. Requests will be taken into consideration, but cannot be guaranteed. The IUHPFL office encourages prospective participants and their parents to be open to placements in any one of our five Spanish-speaking sites.

Step Three: Possible Interviews/Acceptance

- Students will be notified of their semi-final selection status in mid to late February. Semi-final selection statuses include:
    - Acceptance: the student is accepted into the IUHPFL outright based on the strength of his/her application materials.
    - Interview: the student will be interviewed to determine acceptance into the IUHPFL, or the student may be interviewed to determine whether or not they will be granted a spot as an alternate
    - Alternate: the student will be placed on a waiting list as an alternate for the IUHPFL
    - Not accepted: the student is not accepted into the IUHPFL

APPENDIX C: PROFICIENCY TEST

Proficiency test (45 items)

Instructions: Read the story below about a two friends at a party and select the answer that best completes each sentence.

Nombre: _____
Fecha: _____

Selecciona la mejor opción entra las que se ofrecen para cada espacio en blanco.

Creo que es muy interesante _____ de los hábitos alimenticios de la gente. Yo, por mi
(1) a. hablo
b. hablar
c. hablando

parte, _____ vegetariana. Cuando voy a eventos sociales, como por ejemplo fiestas, bodas o
(2) a. soy
b. estoy
c. tengo

bailes, espero que _____ comida vegetariana allí. Algunas personas dicen que _____ supone un
(3) a. hay
b. haya
c. sea
(4) a. le
b. los
c. les

inconveniente proveer _____, pero yo creo que no _____ que ser así. De hecho, la comida
(5) a. lo
b. la
c. le
(6) a. tiene
b. tenga
c. tengo

vegetariana es muy fácil _____ preparar. Y cuando no _____ ofrece, puede ser _____ gran
(7) a. en
b. a
c. de
(8) a. la
b. le
c. se
(9) a. un
b. una
c. el

problema. Yo recuerdo una vez que _____ a una fiesta de cumpleaños y _____ ser todo un
(10) a. fui
b. iba
c. voy
(11) a. resultó
b. resultaba
c. resulté

desastre. La fiesta _____ en la casa de un amigo, y él había invitado a mucha gente. Me
(12) a. estaba
b. era
c. había

sorprendió porque _____ ser un chico sin trabajo, _____ una gran variedad de comida para

(13) a. entre
b. por
c. para

(14) a. tuvo
b. tenía
c. tuviera

para los invitados. Yo creo que si me _____ tocado a mí dar la fiesta, no _____

(15) a. hubiera
b. habría
c. había

(16) a. hubiera
b. habría
c. había

dado ni la mitad de lo que _____ allí. Pero pronto me _____ cuenta que él no había

(17) a. era
b. había
c. hubiera

(18) a. doy
b. daba
c. di

preparado nada vegetariano. Yo no pongo problemas por ese tipo de cosas, pero una amiga

_____ sí _____ hace. _____ a quejarse en frente de todos, mientras el anfitrión

(19) a. mía
b. mi
c. de mí

(20) a. le
b. se
c. lo

(21) a. Empezó
b. Empezaba
c. Empezado

sólo _____ la escena con _____ boca abierta. Yo le dije a mi amiga que _____ de causar tanto

(22) a. miró
b. miraba
c. miraría

(23) a. su
b. una
c. la

(24) a. dejaba
b. deje
c. dejara

escándalo, pero no me puso atención. Por fin, el anfitrión dijo: "La próxima vez que tenga una

fiesta, _____ algo vegetariano." Mi amiga se _____ muchísimo y se _____a poner muy

(25) a. prepararé
b. prepararía
c. preparara

(26) a. enfada
b. enfadó
c. enfadará

(27) a. empezó
b. empieza
c. empezará

nerviosa. Insultó al anfitrión y _____ dijo que _____ muy poco considerado. Yo pensé: "ojalá

(28) a. lo
b. la
c. le

(29) a. fue
b. era
c. estaba

nunca _____ traído a _____ chica aquí". En principio, mi única intención era que ella

(30) a. habría
b. haya
c. hubiera

(31) a. este
b. esta
c. esto

lo _____ bien. Ahora veo que obviamente me _____. Nunca debí _____

(32) a. pasa
b. pasara
c. pase

(33) a. equivoqué
b. equivocaba
c. equivoco

(34) a. haberla
b. haber
c. habido

traído a mi amiga. Yo ya sabía que ella no _____ comportarse adecuadamente.

(35) a. podía
b. puede
c. podría

Antes de dejar la fiesta, le _____ a mi amiga: "¡En cuanto nos _____ ido,

(36) a. dijo
b. decía
c. dije

(37) a. habríamos
b. habremos
c. hayamos

cogeremos un taxi y _____ la fiesta en mi casa donde sí _____ comida

(38) a. seguiremos
b. seguíamos
c. seguiríamos

(39) a. tendríamos
b. teníamos
c. tendremos

vegetariana!". Mi amiga _____ muy contenta y me agradeció que _____ tan buena

(40) a. se pone
b. se puso
c. se ponía

(41) a. soy
b. era
c. fuera

amiga. Yo le dije que se _____ ya y que _____ en toda la comida vegetariana que

(42) a. callaba
b. callaría
c. callara

(43) a. pensara
b. pensará
c. pensaría

_____ a tener en mi casa. Finalmente, ella se _____ y nos fuimos a casa.

(44) a. vamos
b. íbamos
c. fuimos

(45) a. disculpa
b. disculpó
c. disculpaba

APPENDIX D: STIMULUS WORDS USED IN THE IM & AH SPANISH SEMANTIC CLASSIFICATION TASK FOR TIMES 1, 2 & 3

Time 1 warm-up trials: *Asiento, casa, chimpancé, estrella, gallina, mesa, monja, niña, tren, vaca*. (10)

Time 1 experimental trials: *Abeja, abuelo, abrigo, actor, adulto, amigo, anillo, araña, armario, autobús, avión, bailarín, ballena, bañera, bicicleta, bocadillo, bolígrafo, bombilla, botella, caja, calle, cama, camisa, cerdo, cerebro, cesta, cinturón, conejo, copa, corona, cuchara, cuñada, dama, dentista, doctor, dulce, enfermera, esposa, estudiante, falda, foto, galleta, gato, gorra, hermano, hija, hijo, hormiga, iglesia, jarrón, jersey, jirafa, lámpara, libro, lupa, mariposa, medalla, mono, mueble, mujer, nevera, niño, novia, oreja, oso, oveja, padre, pájaro, pan, pantalón, pato, patrón, pavo, perfumen, perro, persona, plato, profesor, puerta, queso, ratón, reloj, rinoceronte, secretaria, serpiente, sofá, tambor, tarta, tiburón, tigre, tijeras, tío, toro, tortuga, vaca, vela, vendedor, vestido, zapato, zorro*. (100)

Animate: *abeja, abuela, actor, adulto, amigo, araña, bailarín, ballena, cerdo, conejo, cuñada, dama, dentista, doctor, enfermera, esposa, estudiante, gato, hermano, hija, hijo, hormiga, jirafa, mariposa, mono, mujer, niño, novia, oso, oveja, padre, pájaro, pato, patrón, pavo, perro, persona, profesor, ratón, rinoceronte, secretaria, serpiente, tiburón, tigre, tío, toro, tortuga, vaca, vendedor, zorro.* (50)

Inanimate: *abrigo, anillo, armario, autobús, avión, bañera, bicicleta, bocadillo, bolígrafo, bombilla, botella, caja, calle, cama, camisa, cerebro, cesta, cinturón, copa, corona, cuchara, dulce, falda, foto, galleta, gorra, iglesia, jarrón, jersey, lámpara, libro, lupa, medalla, mueble, nevera, oreja, pan, pantalón, perfumen, plato, puerta, queso, reloj, sofá, tambor, tarta, tijeras, vela, vestido, zapato.* (50)

Time 2 warm-up trials: *Asiento, casa, chimpancé, estrella, gallina, mesa, monja, niña, tren, vaca*.

Time 2 experimental trials: *Abeja, actriz, adulto, anillo, araña, armario, avenida, bañador, barrio, batidora, bicicleta, bolígrafo, bolso, bragas, burro, caja, calzoncillo, cama, cámara, camarero, carta, cerdo, cerebro, cesta, cirujano, cobaya, conserje, copa, cordero, cruz, cuadro, cucaracha, cuchara, champú, chófer, doctor, elefante, enfermera, enfermo, estudiante, galleta, ganso, gato, gorra, grifo, hija, hipopótamo, hormiga, iglesia, jardín, lámpara, libro, linterna, llave, lobo, lupa, madre, mariposa, mascota, medalla, médico, micrófono, mono, monopatín, mosca, mujer, nevera, niño, novela, novia, oreja, oso, pájaro, paloma, pantalón, pantalla, persona, pez, plancha, plato, primo, profesor, puerta, queso, radio, rana, ratón, revista, sábana, salmón, sombrero, tambor, tenedor, tiburón, tigre, torero, toro, tortuga, vaca, vela*. (100)

Old words from Time 1 (new at Time 1):

Animate: *abeja, adulto, araña, cerdo, doctor, enfermera, estudiante, gato, hija, hormiga, mariposa, mono, mujer, niño, novia, oso, pájaro, persona, profesor, ratón, tiburón, tigre, toro, tortuga, vaca. (25 animate).*

Inanimate: *anillo, armario, bicicleta, bolígrafo, caja, cama, cerebro, cesta, copa, cuchara, galleta, gorra, iglesia, lámpara, libro, lupa, medalla, nevera, oreja, pantalón, plato, puerta, queso, tambor, vela. (25 inanimate).*

New words at Time 2

Animate: *actriz, burro, camarero, cirujano, cobaya, conserje, cordero, cucaracha, chófer, elefante, enfermo, ganso, hipopótamo, lobo, madre, mascota, médico, mosca, c, paloma, pez, primo, rana, salmón, torero. (25 animate).*

Inanimate: *Avenida, bañador, barrio, batidora, bolso, bragas, calzoncillo, carta, cruz, cuadro, champú, grifo, jardín, linterna, llave, micrófono, monopatín, novela, pantalla, plancha, radio, revista, sábana, sombrero, tenedor. (25 inanimate).*

Time 3 warm-up trials: *Asiento, casa, chimpancé, estrella, gallina, mesa, monja, niña, tren, vaca.*

Time 3 experimental trials: *Abrigo, actor, autobús, bailarín, bañador, bañera, baño, barrio, bolso, bombero, botella, búho, caballo, cable, calle, cama, cámara, camarero, cangrejo, carta, chófer, cine, cirujano, cocina, cocinero, conejo, conserje, corona, cruz, cuadro, cucaracha, cuchara, cuchillo, dentista, dependienta, diamante, director, enfermo, entrevistador, espada, gacela, gafas, guerra, helado, hermano, hielo, hormiga, impresora, ingeniero, lata, lavabo, liebre, linterna, limpiadora, lobo, madre, mascota, mono, mosca, motor, naranja, nieto, niño, obrero, oveja, pájaro, paloma, paloma, pan, pantalla, película, peluquero, periodista, perro, playa, presidente, primo, pulsera, rana, rana, rata, refresco, rinoceronte, sábana, serpiente, sofá, tarta, televisor, tenedor, toalla, torero, torero, traductor, turista, ventilador, vestido, vino, yerno, yogurt, zapato.* (100)

Old words from Time 1 (new at Time 1):

Animate: *Actor, bailarín, dentista, hermano, hormiga, niño, oveja, pájaro, perro, rinoceronte, serpiente, mono*. (12 animate)

Inanimate: *Abrigo, autobús, bañera, botella, calle, cama, corona, cuchara, pan, sofá, tarta, vestido, zapato*. (13 inanimate)

Old words from Time 2 (new at Time 2):

Animate: *Camarero, conserje, chófer, lobo, madre, mascota, mosca, paloma, primo, torero, rana, paloma, cirujano.* (13 animate)

Inanimate: *Bañador, barrio, bolso, carta, cámara, cruz, cuadro, cucaracha, linterna, pantalla, sábana, tenedor.* (12 inanimate)

New words at Time 3

Animate: *Bombero, búho, caballo, cangrejo, cocinero, conejo, dependienta, diamante, director, enfermo, entrevistador, gacela, ingeniero, liebre, nieto, obrero, peluquero, periodista, presidente, rana, rata, torero, traductor, turista, yerno*. (25 animate)

Inanimate: *baño, cable, cine, cocina, cuchillo, espada, gafas, guerra, helado, hielo, impresora, lata, lavabo, limpiadora, motor, naranja, película, playa, pulsera, refresco, televisor, toalla, ventilador, vino, yogurt*. (25 inanimate)

APPENDIX E: Stimulus words used in the IM & AH English semantic classification task for Times 1 & 2 (same task at both times).

Time 1&2 warm-up trials: *father, janitor, leopard, sheep, sink, table, tree, wallet, window, zebra*. (10)

Time 1&2 experimental trials: *Actor, adult, ant, basket, bear, bee, belt, bench, bicycle, bird, blanket, board, boat, book, boy, bride, brother, building, car, cat, chair, child, chimney, clothes, comb, cow, dancer, daughter, dentist, desk, dictionary, doctor, dog, door, duck, farmer, father, female, fireplace, floor, garbage, girl, goat, horse, human, husband, ink, janitor, job, judge, key, king, kitchen, knife, lady, lamp, leopard, luggage, maple, medal, monkey, mother, mouse, nephew, newspaper, nurse, oak, pants, parent, pencil, perfume, person, picture, pig, pillow, plane, rabbit, rat, road, roof, secretary, sheep, ship, sink, snake, son, spoon, stamp, step, street, student, table, tape, teacher, television, tiger, tire, tree, trophy, truck, tunnel, turtle, umbrella, wall, wallet, wife, window, wolf, woman, zebra*. (110)

Animate: *Actor, adult, ant, bear, bee, bird, boy, bride, brother, cat, child, cow, dancer, daughter, dentist, doctor, dog, duck, farmer, father, female, girl, goat, horse, human, husband, janitor, judge, king, lady, leopard, maple, monkey, mother, mouse, nephew, nurse, oak, parent, person, pig, rabbit, rat, secretary, sheep, snake, son, student, teacher, tiger, turtle, wife, wolf, woman, zebra.* (55 animate)

Inanimate: ***Basket, belt, bench, bicycle, blanket, board, boat, book, building, car, chair, chimney, clothes, comb, desk, dictionary, door, fireplace, floor, garbage, ink, job, key, kitchen, knife, lamp, luggage, medal, newspaper, pants, pencil, perfume, picture, pillow, plane, road, roof, ship, sink, spoon, stamp, step, street, table, tape, television, tire, tree, trophy, truck, tunnel, umbrella, wall, wallet, window*** **(55 inanimate)**

APPENDIX F: Stimulus words used in the IM & AH picture-naming task for Times 1, 2 & 3.

Picture-naming task (Time1):

Time 1 warm-up trials: *abrigo (coat), águila (eagle), ángel (angel), uva (grape), tunel (tunnel), flor (flower), columna (column), nube (cloud) uña (nail), violín (violin).* **(10)**

Time 1 experimental trials: *Ambulancia (ambulance), anillo (ring), araña (spider), árbol (tree), arco (arc), arco (arch), ardilla (squirrel), armadillo (armadillo), armario (closet), armónica (harmonic), arpa (harp), aspiradora (vacuum-cleaner), astronauta (astronaut), autobús (bus), avión (airplane), bailarina (dancer), balcón (balcony), banco (bench), bandera (flag), bañera (bathtub), bicicleta (bicycle), cebolla (onion), diamante (diamond), diente (tooth), dinosaurio (dinosaur), dragón (dragon), elefante (elephant), enfermera (nurse), escorpión (scorpion), escritorio (desk), espaguetis (spaghetti), esqueleto (skeleton), estrella (star), flauta (flute), flecha (arrow), fresa (strawberry), gafas (glasses), gallina (chicken), gato (cat), goma (eraser), gorila (gorilla), gorra (cap), gota (drop), guitarra (guitar), hamburguesa (hamburger), helicóptero (helicopter), huevo (egg), iglesia (church), jirafa (giraffe), labios (lips), lámpara (lamp), libro (book), llave (key), luna (moon), maleta (suitcase), mano (hand), manzana (apple), micrófono (microphone), mono (monkey), montana (mountain), mosca (fly), niño (boy), oveja (sheep), pájaro (bird), pantalones (pants), paraguas (umbrella), payaso (clown), tren (train), periódico (newspaper), perro (dog), pez (fish), piano (piano), pie (foot), pierna (leg), planeta (planeta), pollo (chicken), queso (cheese), regalo (present), regla (ruler), reina (queen), sangre (blood), tren (train), sofá (couch), tiburón (shark), tijeras (scissors), toro (bull), tortuga (turtle), tractor (tractor), tren (train), trompeta (trumpet), vaca (cow), vagón (truck), vaso (glass), vela (candle), ventana (window), ventilador (fan), vestido (dress), volcán (volcano), zapato (shoe), zorro (fox).* **(100)**

Time 2 warm-up trials: *Abrigo (coat), águila (eagle), ángel (angel), uva (grape), tunel (tunnel), flor (flower), columna (column), nube (cloud) uña (nail), violín (violin).* **(10)**

Time 2 experimental trials: *Almohada (pillow), araña (spider), arco (arc), ardilla (squirrel), armario (closet), arpa (harp), astronauta (astronaut), avión (airplane), balcón (balcony), bañador (swimming-suit), bandera (flag), barco (boat), basura (garbage-can), bicicleta (bicycle), bigote (mustache), blusa (blouse), bombillas (lightbulb), búfalo (buffalo), burro (donkey), caballo (horse), cadena (necklace), calculadora (calculator), cama (bed), camello (camel), carro (car), carta (letter), casa (house), castillo (castle), cebra (zebra), cerebro (brain), chica (girl), cigarrillo (cigarette), cocina (kitchen), cuadro (picture), dedo (finger), diamante (diamond), dinosaurio (dinosaur), disco (record), elefante (elephant), escalera (ladder), escorpión (scorpion), espada (sword), espaguetis (spaghetti), estrella (star), flecha (arrow), fresa (strawberry), gallina (chicken), globo (balloon), goma (eraser), gorila (gorilla), gorra (cap), guitarra (guitar), helicóptero (helicopter), iglesia (church), inodoro (lavatory), labios (lips), lápiz (pencil), libro (book), luna (moon), mano (hand), manzana (apple), mariposa (butterfly), martillo (hammer), melocotón (peach), mesa (table), micrófono (microphone), moneda (coin), montaña (mountain), mujer (woman), mundo (world), naranja (orange), nariz (nose), niño (boy), ojo (eye), oveja (sheep), pan (bread), pantalones (pants), payaso (clown),*

pelota (ball), perro (dog), piano (piano), pierna (leg), piscina (swimming-pool), planeta (planet), plato (dish), pollo (chicken), puente (bridge), queso (cheese), regla (ruler), reloj (watch), sangre (blood), sofá (couch), tijeras (scissors), tortuga (turtle), tren (train), vagón (truck), vela (candle), ventilador (fan), zanahoria (carrot), zapato (shoe). **(100)**

Old words from Time 1: *Araña (spider), arco (arc), ardilla (squirrel), armario (closet), arpa (harp), astronauta (astronaut), avión (airplane), balcón (balcony), bandera (flag), bicicleta (bicycle), diamante (diamond), dinosaurio (dinosaur), elefante (elephant), escorpión (scorpion), espaguetis (spaghetti), estrella (star), flecha (arrow), fresa (strawberry), gallina (chicken), goma (eraser), gorila (gorilla), gorra (cap), guitarra (guitar), helicóptero (helicopter), iglesia (church), labios (lips), libro (book), luna (moon), mano (hand), manzana (apple), micrófono (microphone), montaña (mountain), niño (boy), oveja (sheep), payaso (clown), perro (dog), piano (piano), planeta (planet), pollo (chicken), queso (cheese), regla (ruler), sangre (blood), sofá (couch), tijeras (scissors), tortuga (turtle), tren (train), vagón (truck), vela (candle), ventilador (fan), zapato (shoe).* (50)

New words at Time 2: *Almohada (pillow), bañador (swimming-suit), barco (boat), basura (garbage-can), bigote (mustache), blusa (blouse), bombillas (lightbulb), búfalo (buffalo), burro (donkey), caballo (horse), cadena (necklace), calculadora (calculator), cama (bed), camello (camel), carro (car), carta (letter), casa (house), castillo (castle), cebra (zebra), cerebro (brain), chica (girl), chimenea (chimney), cigarrillo (cigarette), cocina (kitchen), cuadro (picture), dedo (finger), disco (record), escalera (ladder), espada (sword), globo (balloon), inodoro (lavatory), lápiz (pencil), mariposa (butterfly), martillo (hammer), melocotón (peach), mesa (table), moneda (coin), mujer (woman), mundo (world), naranja (orange), nariz (nose), ojo (eye), pan (bread), pelota (ball), pierna (leg), piscina (swimming-pool), plato (dish), puente (bridge), reloj (watch), zanahoria (carrot).* (50)

Time 3 warm-up trials: *Abrigo (coat), águila (eagle), ángel (angel), uva (grape), tunel (tunnel), flor (flower), columna (column), nube (cloud) uña (nail), violín (violin).* **(10)**

Time 3 experimental trials: *Ambulancia (ambulance), árbol (tree), armónica (harmonic), autobús (bus), ballena (whale), bañador (swimming-suit), banco (bench), basura (garbage-can), blusa (blouse), bomba (bomb), botella (bottle), búfalo (buffalo), bufanda (scarf), caballo (horse), calculadora (calculator), camello (camel), campana (bell), carretera (road), carta (letter), castillo (castle), cebolla (onion), cerebro (brain), chica (girl), cisne (swan), cocina (kitchen), cruz (cross), cuchillo (knife), dedo (finger), dragón (dragon), escalera (ladder), escritorio (desk), espejo (mirror), flauta (flute), gafas (glasses), globo (balloon), gorila (gorilla), hamburguesa (hamburger), hoja (leaf), jirafa (giraffe), ladrillo (brick), lápiz (pencil), león (lion), limón (lemon), llave (key), loro (parrot), mago (magician), manzana (apple), mar (sea), martillo (hammer), mesa (table), molino (windmill), mosca (fly), motocicleta (motorcycle), mujer (woman), naranja (orange), nevera (refrigerator), oreja (ear), pájaro (bird), palmera (palmtree), pan (bread), pastel (cake), pato (duck), pecera (fishbowl), pelo (hair), pera (pear), pez (fish), pimiento (pepper), piscina (swimming-pool), plátano (banana), pollo (chicken), pozo (well), pulmones (lungs), queso (cheese), regalo (present), revista (magazine), rueda (wheel), sandía (watermelon), sartén (frying-pan), semáforo (traffic-light), serpiente (snake), silla (chair), sobre (envelope), sol (sun), sombrero (hat), tacón (heel), taza (cup), teclado (keyboard),*

tejado (roof), teléfono (telephone), televisión (television), tenedor (fork), termómetro (thermometer), tigre (tiger), tomate (tomato), toro (bull), tostador (toaster), trompeta (trumpet), vaso (glass), vestido (dress), zanahoria (carrot). **(100)**

Old words from Time 1: *Autobús (bus), ambulancia (ambulance), árbol (tree), armónica (harmonic), banco (bench), cebolla (onion), dragón (dragon), escritorio (desk), flauta (flute), gafas (glasses), gorila (gorilla), hamburguesa (hamburger), jirafa (giraffe), llave (key), manzana (apple), mosca (fly), pájaro (bird), pelo (hair), pez (fish), regalo (present), silla (chair), toro (bull), trompeta (trumpet), vaso (glass), vestido (dress).* **(50)**

Old words from Time 2: *Bañador (swimming-suit), basura (garbage-can), blusa (blouse), búfalo (buffalo), caballo (horse), calculadora (calculator), camello (camel), carta (letter), castillo (castle), cerebro (brain), chica (girl), cocina (kitchen), dedo (finger), escalera (ladder), globo (balloon), lápiz (pencil), martillo (hammer), mesa (table), mujer (woman), naranja (orange), pan (bread), piscina (swimming-pool), pollo (chicken), queso (cheese), zanahoria (carrot).* **(50)**

New words at Time 3: *Ballena (whale), bomba (bomb), botella (bottle), bufanda (scarf), campana (bell), carretera (road), cisne (swan), cruz (cross), cuchillo (knife), espejo (mirror), hoja (leaf), ladrillo (brick), león (lion), limón (lemon), loro (parrot), mago (magician), mar (sea), molino (windmill), motocicleta (motorcycle), nevera (refrigerator), oreja (ear), palmera (palmtree), pastel (cake), pato (duck), pecera (fishbowl), pera (pear), pimiento (pepper), plátano (banana), pozo (well), pulmones (lungs), revista (magazine), rueda (wheel), sandía (watermelon), sartén (frying-pan), semáforo (traffic-light), serpiente (snake), sobre (envelope), sol (sun), sombrero (hat), tacón (heel), taza (cup), teclado (keyboard), tejado (roof), teléfono (telephone), televisión (television), tenedor (fork), termómetro (thermometer), tigre (tiger), tomate (tomato), tostador (toaster).* **(100)**