THIS IS AN EARLIER VERSION OF THE MANUSCRIPT. FOR THE FINAL VERSION, PLEASE CHECK THE JOURNAL WEBSITE:
Serrano, R., Llanes, À.; Tragant, E. (2016). Examining L2 development in two short-term intensive programs: Study abroad vs. "at home". *System*, *57*, 43-54.

Examining L2 development in two short-term intensive programs for teenagers: study abroad vs. “at home”

Abstract

Although research on the effect of spending time abroad on language learning has become popular in the last few years, there are only a few studies that concentrate on short-term stays for teenagers or that include intensive second language (L2) programs as their comparison groups. The purpose of this study is to fill these gaps and analyze the L2 development of two groups of Catalan/Spanish bilingual learners of English in two short-term programs: a study abroad (SA) in the UK (n= 54) and an intensive course “at home” (AH) (n= 58). The learners (mean age= 14.37) were tested on several tasks that focused on different language areas (grammar, formulaic sequences, written and oral production) before and after their respective program. The results of our analyses suggest that the participants in the two contexts significantly improved in most of the measures. Our findings also indicate that the SA context was more beneficial for the improvement of oral lexical richness and the learners in the AH context were slightly better in their receptive knowledge of grammar. These results seem to be the product of the specificity of the two contexts and the type and amount of L2 use that they offer.

Keywords: learning context, study abroad, intensive courses, short-term L2 programs.

1. Introduction

In the last few years there has been an increasing interest in the Second Language Acquisition (SLA) literature concerning the effect of learning context on second language (L2) development (Housen, 2012; Housen et al., 2011). One of the contexts that has received more attention is the "study abroad" context, which offers a combination of L2 instruction in a rather formal setting (i.e., classes) with immersion in the native speech community (Segalowitz & Freed, 2004). Other terms used in the literature are "residence abroad" or "stay abroad", which refer to those programs that do not necessarily include the L2 instruction component. In this section the relevant literature on learning context will be reviewed, focusing on the following aspects: L2 development, L2 use, intensity and short-term programs for teenagers.

1.1. L2 development in the SA context

Researchers have been trying to empirically assess the "folk myth" that the best way to learn a foreign language is to spend some time abroad. The studies so far have shown some results that back up this belief, but not for all the areas under examination. Concerning oral fluency, most studies suggest that this area typically improves after spending some time abroad (no matter how short the time is) (DeKeyser, 2014; Llanes,

2011). For example, Segalowitz and Freed (2004) reported that, after a semester abroad, a group of North-American college students (n= 22) significantly improved in several measures of oral fluency in Spanish; namely, speech rate, number of words in the longest turn, longest fluent run, and absence of filled pauses. Likewise, Mora and Valls-Ferrer (2012) also reported a significantly positive effect of SA on oral fluency, this time for 30 college students who were learning English as an L2. As for vocabulary development, most of the studies show that participants increase their lexical repertoire as a result of an SA experience (Dewey, 2008; Foster, 2009; Ife et al., 2000). On the other hand, there are other L2 areas, such as grammar, for which the SA context does not appear to be so beneficial. For instance, Collentine (2004) found that a group of 26 North-American college students did not improve their Spanish grammar skills significantly after spending one semester in Alicante, Spain. Similarly, in the case of pragmatics, the literature shows contradictory findings. Taguchi (2015) suggests that the learners abroad do not always improve their pragmatic/sociolinguistic skills and that there is a "shaky link between study abroad experience and pragmatic gains" (p. 16).

1.2. L2 use

Most studies that have examined L2 development in the study abroad (SA) context have often compared this context with classroom instruction in the participants' home country (or at home, AH, context). Comparing the two settings, variables such as the quantity and quality of L2 input as well as the opportunities of producing L2 output and engaging in interaction in the L2 are all potentially higher in the SA context (see also Sanz, 2014).

The significant roles of these factors in L2 learning in general has already been established, and it is interesting to examine the ways in which learners benefit from a context (SA) where L2 input and interaction are more easily accessible.

Nevertheless, despite the vast number of opportunities for L2 practice that the SA context theoretically offers, research has shown that not all learners benefit from these opportunities equally. Ranta and Meckelborg (2013) report that some learners use the L2 significantly more than others, and that some are more inclined to practice their productive skills and others their receptive skills (or, similarly, some prefer aural to written input or the other way round). In the same way, Kinginger (2008) demonstrates that some SA participants hardly ever use the L2 and therefore they experience few (if any) L2 gains: out of the 24 participants in her study, 13 did not experience any change in their L2 level. However, there are some learners that integrate within the L2 society, which offers them multiple opportunities to practice their L2 skills, and thereby they significantly improve their L2 (three out of the 24 students in Kinginger's study moved from the basic level to the advanced one).

The range of opportunities to use the L2 abroad also varies depending on SA program type. While there are some SA programs that encourage participants to use the L2 to a high degree, there are others in which this is not the case. An example of the former type would be the program for child L2 learners described in Llanes and Muñoz (2013). In that program, a group of Spanish-Catalan bilingual learners of English spent three months in Ireland, living with Irish families and attending regular Irish schools. In both contexts, they were forced to interact in English, as every child was placed in a home and in a class where there were no other Spanish or Catalan speakers. An

illustration of an SA experience that would not encourage so much L2 use would be the example of a university exchange student living in a residence hall or in an apartment together with other L1-speaking students (as was true of some adult learners in that same study).

1.3. Intensity

The differences between the SA and AH context can also be analyzed in terms of "intensity" of language exposure, another well-established factor in SLA. The SA context typically provides not only more contact hours than the typical AH context but also more intensive exposure to the L2. When analyzing L2 acquisition in general, Carroll (1967) suggests that more time learning the target language usually translates into higher proficiency levels (all other things being equal). However, Stern (1985) claims that, when analyzing time, it is important to consider not only "total time" (that is, duration) but also time distribution. Stern suggests that concentrated exposure to the L2 might be more beneficial for learning than spreading out the contact hours over long periods of time. Several studies have confirmed the positive effect of intensive exposure to the L2 in classroom settings in contrast to the "drip-feed" approach that characterizes many EFL programs (Spada & Lightbown, 1989). For instance, White and Turner (2005) found that L1 French-speaking children receiving intensive English instruction in Canada (approximately 400 hours in five months) significantly outperformed their peers receiving regular instruction (36 hours in 10 months) in oral production in several tasks. As for adults, Serrano (2011) found that a group of Spanish learners with intermediate

English L2 proficiency enrolled in a one-month intensive course improved certain L2 skills more significantly than a comparison group who received the same number of hours of instruction (around 80) distributed over a whole academic year.

As the SA setting provides more intensive exposure to the L2 than the typical AH setting, more L2 gains can also be expected, following Stern's argument. The previous studies that have found an advantage for the SA context included "regular" (or non-intensive) AH programs, which typically provide 2-4 hours/week of instruction. However, the studies that have compared the SA context with programs that offer intensive language contact (e.g., "domestic immersion" or intensive AH courses) have not found advantages for the SA setting for adult learners.

Freed, Segalowitz, and Dewey (2004) compared the oral fluency gains in L2 French by a group of US college students learning the L2 in three different settings: SA (n= 8), "domestic immersion" (IM) (n= 12), and AH (n= 8). The term "domestic immersion" refers to a context which combines both classroom and out-of-class L2 exposure, but in the learners' home country (without access to a native speech community). In the study by Freed and colleagues, the L2 learners stayed in the US and lived in dormitories together with other French learners. In order to ensure L2 use, all the participants pledged to speak only French during the 7-week program. Additionally, the learners received 17.5 hours of classroom instruction per week. The authors found that while the SA setting was clearly superior to the AH setting to improve the participants' oral fluency, it was the participants in the IM context who experienced the greatest gains. The authors suggest that the higher number of L2 contact hours the participants in the IM setting received accounted for their superior performance.

In a similar vein, Serrano, Llanes, and Tragant (2011) compared the L2 development of a group of Catalan/Spanish learners of English in three different settings: SA (n= 25), intensive course AH (n= 69), and semi-intensive course AH (n= 37). In contrast to the "domestic immersion" program, the learners in the two AH courses were exposed to the L2 only in class (for 25 hours/week in the intensive course and 10 hours/week in the semi-intensive). The authors found that, whereas the SA setting was superior to the semi-intensive domestic course, no differences were found between the SA program and the intensive course AH considering several measures of written and oral production. This seems to indicate that intensive L2 language contact is beneficial for L2 development, whether at home or abroad.

1.4. Short-term programs for teenagers

The present study aims to delve deeper into the differences between learning contexts by considering two types of intensive programs, one abroad and the other one at home, in order to throw more light on the issue of intensive exposure. We have chosen two programs that are quite short (3-4-week summer programs) because it is interesting to examine whether such a short time would be enough for learners to improve certain aspects of the L2 significantly. There are a few studies that have confirmed that short SA programs can lead to significant improvement of L2 skills in the case of adult learners (Allen & Herron, 2003; Cubillos, Chieffo, & Fan, 2008; Rodrigo, 2011). However, short-term SA programs are under-researched in comparison with stays of one semester (or more), though they are very popular, especially among teenagers. This is one reason why

we have included this population in our study. Additionally, it is important to investigate how their L2 develops in an SA context in comparison to other age groups.

Previous research on age and L2 acquisition in the SA context has focused on children and adults. The results from these studies show that the SA context seems to be more beneficial for children (Llanes & Muñoz, 2013 and Muñoz & Llanes, 2014). This finding is, one the one hand, attributed to cognitive differences: children may make more use of (and thus benefit more from) their implicit learning mechanisms than adults. On the other hand, there are also sociological differences associated with age: age determines certain aspects related to L2 use, such as the type of interlocutors, type of accommodation (to mention a few), and these factors play a crucial role when learning an L2 in an SA context. Including adolescents in research studies is important because they are "in between" children and adults concerning aspects of their L2 learning experience. For example, their age allows them to choose the types of interlocutors to a greater extent than children and also move more freely in the SA context (e.g., they can go shopping by themselves, they can order drinks or food, they can buy a ticket for the movies, etc.), but at the same time their movements are more restricted than those of adults.

The few studies that have analyzed adolescents' L2 development in an SA context suggest that this population benefits from short-term stays. Evans and Fisher (2005) investigated the effects of 6-11 days in France by a group of British teenagers (n= 69) and found significant gains in the standardized language French test after such a short period abroad. Similarly, Llanes and Muñoz (2009) also focused on the effects of a short SA experience on a group of Spanish-Catalan bilingual teenagers (n= 24). The authors investigated the impact of a 3-4-week stay in the UK on the participants' oral fluency and

listening comprehension skills and found that such SA programs have a positive impact on the development of the participants' L2 oral and listening skills.

The current study aims to further contribute to the literature on the effect of short-term SA programs for teenagers by incorporating some features that were not included in the two previously-mentioned studies that have examined this context for this specific age group. First, we include a comparison group (as in some cases gains between the pre- and posttest could be partly due to task repetition effects). Moreover, including a comparison group allows us the possibility to assess not only the potential advantages of the SA context but also of the AH context. The AH context we have considered is also different from most studies, as it offers intensive instruction and encourages L2 use for an extensive amount of time (which makes it more comparable to the SA setting than the typical AH program usually included as a comparison group in the studies analyzing the effect of the SA context). Additionally, we have chosen a variety of tasks that tap into different L2 areas in order to capture potentially differential effects of learning context on L2 development: a grammaticality judgment test, a formulaic sequences test, an oral narrative (analyzed through a variety of measures), and an elicited imitation task.

The specific questions that guide this study –considering the gaps in the literature identified above– are the following: 1) Do teenage L2 learners significantly improve their L2 skills after a short intensive summer program, whether abroad or at home? 2) Are there differences between the two learning contexts? And if so, are those differences related to the L2 area that is being assessed?

2. Method

2.1. Programs

The two programs that we have considered for this study are a study abroad program in the south of England and an intensive summer program in Barcelona. The programs were organized by two different private institutions and attracted a similar population: these institutions are both located in an area of Barcelona (Spain) where middle and upper-middle class families live.

The SA program took place during the first three weeks of July. The participants in this program were provided with accommodation in a residence hall together with other international students. The program included three hours of classroom instruction each weekday (Monday-Friday) in the morning, as well as sports and other leisure activities in the afternoon. The participants were assigned to different groups for L2 instruction depending on their proficiency level, and they were mixed with other international students. The classes had special focus on listening and speaking activities, through songs, games, etc., but some time was also devoted to grammar and, to a greater extent, vocabulary (for which teachers followed a book especially designed for the SA program). These classes were taught by native English-speaking teachers. After the L2 classes, the participants had lunch in the cafeteria, where they could sit with other international students (although very few actually did). In the afternoon, some of the programmed activities included basketball, hockey, or arts and crafts. These activities were monitored by qualified native English speakers. On weekends, the students traveled to different nearby cities with their L1-speaking peers (therefore these trips were more

enriching from a cultural point of view than from a linguistic perspective) and they could also attend several activities in the evening, such as movies, quizzes, etc. where they met other international students.

Regarding the AH context, the program lasted for four weeks, from the last week of June until the last week of July and included 4.5-hour daily English lessons from Monday through Friday. The students had two breaks of 10 minutes and one of 20 minutes. The students were grouped in terms of their proficiency and two different teachers were in charge of each class: one taught the first session before the long break and the other the second one. The teachers were all qualified English as a Foreign Language (EFL) teachers and they were all from the UK. The courses followed popular EFL textbooks, which, despite including many communicative activities, also focused on language forms. Apart from the textbook, the teachers also used other resources that helped the students practice their speaking and listening skills in ways that were motivating for them: karaoke songs, conversations about pop culture, etc. Additionally, the students went to the computer room quite often and they played computer games in English, following the instructions of their teachers.

2.2. Participants

The total number of participants in this study includes 102 English learners from Catalonia (Spain), who were Spanish/Catalan bilingual. One group of learners (N=54) were in the SA context and another one (N=58) did the intensive course AH. The age for all the participants ranged from 13 to 17. The mean age for the SA context was 14.57 and

for the AH 14.17. To see the age distribution by context, see Table 1. In terms of gender, there were more girls than boys abroad (39 vs. 15) while the opposite was true in the intensive course (18 girls and 40 boys). The learners from the two contexts came from well-educated families as at least 80% of the students had parents who had university degrees (85% in the SA context and 95% in the AH context).

[Table 1]

2.3. Instruments

L2 development was assessed through a variety of tasks in order to have a more general picture of what the learners could do in English at the beginning and at the end of their programs. Additionally, different L2 measures were used with the aim of examining whether there are certain areas of the L2 for which one particular context is especially beneficial. The battery of tests designed for this study included the following tasks: grammaticality judgment test, test of formulaic sequences, oral narrative and elicited imitation task. Each of these tasks will be discussed in turn.

2.3.1. Grammaticality judgment test (GJT)

The main purpose of this test was to examine participants' explicit knowledge of English grammar. Although both contexts (SA and AH) provided L2 instruction, there were more hours of L2 class in the AH setting, and the type of instruction was more focused on

grammar than in the SA setting. Consequently, certain advantages were expected for the AH learners.

This test consisted of 14 items, 11 of which were incorrect and 3 correct. This instrument had good internal consistency (Cronbach alpha coefficient was .700). The GJT was previously piloted with a population that was similar to the one included in the present study. The original test had 20 items (15 incorrect, 5 correct); however, several items had to be eliminated because they proved to be either too difficult or too easy, with little discriminatory power, according to an item analysis. We included more incorrect than correct items because they are less prone to guessing on the part of the learners. In fact, we observed in the pilot test that many students just picked "correct" for most sentences, especially those who were not motivated to do the test or think about possible errors. We consider that including many correct items can therefore lead to an overestimation of participants' grammar knowledge (which might be especially the case for young learners).

The sentences included in the test targeted grammar points that are usually dealt with in EFL books in Spain for teenage learners. Some of the grammar points that we considered include irregular plurals (*foot-feet; child-children*), irregular past (*swim-swam*), word order adverb-verb (*I always drink water*), noun-verb agreement (*George doesn't like water*), comparatives (*more dangerous*), and verb tenses. The learners had to read the sentences and tick one of the three columns that appeared next to it: "correct", "incorrect", or "I don't know". Additionally, the students were asked to correct those sentences they believed to be incorrect. In terms of scoring, we gave participants one point for correct sentences identified as correct, and incorrect sentences successfully

corrected. Other responses were scored as 0. The test had a maximum possible score of 14 points.

2.3.2. Formulaic sequences test (FST)

The purpose of this test was to examine learners' knowledge of formulaic sequences, or typical expressions used in everyday English. While in the GJT we expected the AH learners to show an advantage, for the FST the expectation was that learners in the SA setting would outperform their peers at home, as they can practice the language in an environment where a lot of formulaic expressions (e.g., *how are you today? How much is this?*) are likely to occur naturally. Although there are very few studies that examine how spending time abroad affects the acquisition of FSs, the research that exists in this respect seems to suggest that the SA context contributes significantly to the acquisition of L2 formulas (Taguchi, 2013), especially there is intensity of interaction in the L2 (Bardovi-Harlig & Bastos, 2011). Most of the target formulaic expressions were taken from Nattinger and DeCarrico (1992) and include formulas like *Good luck! Nice to meet you!* or *Take care!* The test consisted of 16 different sentences: 13 were incorrect and 3 correct. The reliability of this instrument was also high, with a Cronbach alpha coefficient reported of .808. As with the GJT, this test was previously piloted. The original test included 20 sentences, 15 of which were incorrect and 5 correct. The item analysis showed that there were 4 sentences that had little discriminatory power and were thus eliminated.

In this test the learners also had to indicate whether the sentences included in the task were correct or incorrect, or alternatively mark “I don’t know”. The learners were asked to correct those sentences they believed to be incorrect. The scoring of this test was the same as for the GJT: one point was given only in the case of correct sentences identified as such, or incorrect sentences successfully corrected. The maximum possible score was 16.

2.3.3. Oral narrative

The task which was used to examine participants’ oral production was an oral narrative based on a series of pictures, called “The Picnic Story” (Heaton, 1966). This story has been repeatedly used for research purposes (Muñoz, 2006) and shows two children preparing a picnic with their mom. While they are busy with the arrangements, their puppy gets into their picnic basket. When they are ready to eat their food, once they are on a little hill, they discover that there is no food left in their basket because their puppy has eaten everything.

The oral narrative was analyzed in terms of fluency, lexical richness, syntactic complexity and accuracy. Fluency was assessed through the number of syllables per minute the L2 learners produced. This count was based on pruned speech, and thus did not include repetitions, false starts, invented words or words in the participants’ first language. Guiraud’s Index (Van Hout & Vermeer, 2007; Vermeer, 2000) was used to measure lexical richness. This index is computed by dividing the word types included in a narrative by the square root of the word tokens. The number of clauses per T-unit was

considered the index of syntactic complexity, also following previous studies (see Bulté & Housen, 2012 for the use of this measure in research as well as other measures of complexity)[1]. Finally, for accuracy we decided to focus on lexical accuracy of nouns and verbs. We established a percentage of correct nouns and verbs, considering the total number of nouns and verbs a particular student used. Additionally, we examined morphosyntactic accuracy by focusing on verbs and examining errors of tense and agreement per T-unit. Through these two measures we hoped to obtain information about accuracy in these two specific target areas[2].

2.3.4. Elicited imitation task

This task consisted of 20 sentences related to a story the students were made familiar with before the task ("The ball story", Heaton, 1966). In this story there were some boys playing with a ball that went into a hole and they had to figure out ways of getting the ball back. The students were asked to repeat (as closely as they could) the sentences that were produced by a native English speaker. As reported in Tracy-Ventura et al. (2013), elicited imitation tasks have been previously used as measures of proficiency by analyzing how successful learners were at repeating the target sentences (Bowden, 2007; Ortega, 2000). Similarly, Gallimore and Tharp (1981) suggest that learners' performance in this type of task correlates with other measures of their L2 performance. Therefore, we used this instrument as another measure of learners' L2 performance.

The sentences included in this study varied in length, but they all ranged between 4 and 10 syllables. Between a given sentence and the prompt for repetition, there was a

pause of 0.5 seconds and a beep. The students then had six seconds to repeat the target sentence. This task was scored considering the number of words the students were able to repeat successfully out of the total number of words (119), following the repetition accuracy measure used in Trofimovich, Lightbown, Halter and Song (2009). The internal consistency of the elicited imitation task was high, with a Cronbach alpha coefficient of .913.

2.3.5. Questionnaires and observations

It is hard to make quantitative analyses of amount and type of input, output and interaction in the two programs. Some attempts have been made in the literature at quantifying L2 learners' hours of exposure and language use, especially in the SA context (e.g., Language Contact Profile, Freed et al., 2004) by asking learners to report on how many hours they spoke, listened to, wrote, read, etc. in the target language. However, the reliability of such instruments is questionable considering the difficulty anyone would experience in thinking about the hours they spent reading, writing, etc. in a particular language in a particular day/week. In the AH setting, the hours the students are exposed to the L2 are easier to estimate for classroom exposure, but the problem comes again when we want to learn about out-of-class exposure. In any case, researchers need to rely either on self-reported questionnaires or observations. We used both to try to investigate how often, with whom, and in which way our participants used English. Instead of asking them for hours of exposure, we used frequency scales and we asked slightly different questions in the two programs, since the two contexts require different L2 use. For the

SA learners, we asked about the frequency of L2 use (*almost never, few times, sometimes, often, almost always*) in different contexts: English class, dining room, afternoon activities, residence hall, weekends and free time. Additionally, we inquired about the nationality of their two closest friends while abroad, and the two adults they interacted with the most, and also the language they normally used in their interactions. We consider that it is easier for students to respond to these questions than to count the hours of speaking, listening, etc. in the L2, as in the Language Contact Profile. The answers to the questions we proposed are less quantifiable, but they are probably more reliable (see McManus, Mitchell, & Tracy-Ventura, 2014 for more in-depth methods to obtain information about SA learners' social networks). For the learners AH, we asked about the use of English in class and during breaks with teachers and classmates, and the frequency of out-of-class reading, writing, speaking, listening in English, as well as grammar and vocabulary practice. Different questions were included for the SA and for the AH learners due to the intrinsic differences between the two contexts concerning the opportunities for L2 use.

Apart from the self-reported questionnaires, we have data from our own observations, as we spent one day observing the learners in the two contexts. In the SA context, we observed the English classes, the breaks during the classes, had lunch and dinner with the students in the cafeteria, and spent an afternoon with the learners and their group leaders. For the AH context, we observed the classes and the breaks. The results of the questionnaires and the observations will be included in the discussion to help explain the results.

2.4. Procedure

The design of this study is pretest-posttest in order for us to examine language gains. The pretest was administered at the beginning of each program and the posttest two days before the end: there were a total of 18 days of L2 exposure (including every day of the week in the SA context and the days of class in the AH context) between the two testing times. The learners performed the same tasks at the two testing points, except for the questionnaire, which was performed only in the posttest.

The tasks were administered in classrooms in the context where the participants were (abroad and "at home") and instructions were always provided in the students' first language (L1), to make sure they understood what they had to do. First, the students performed the GJT (7 minutes) and then the FST (5 minutes). After these tasks, the learners filled out the questionnaire, which was also administered in their L1. For the oral narrative and the elicited imitation task, the students were taken out of their classes into a room with the researcher and did the tasks individually: first, the oral narrative, and then the elicited imitation task. Because of practical reasons related to data collection, only a subgroup of participants performed these two tasks (SA= 16, AH= 24). For the oral narrative, the participants were shown the pictures for approximately 30 seconds until they were familiar with them and thus ready to tell the story. The learners typically took one and a half to two minutes to narrate the story. For the elicited imitation task, the total duration was 10 minutes. The students listened to the instructions and the sentences on a CD. They were also shown the pictures the sentences referred to for 30 seconds, and then they had some practice sentences before the 20 target sentences appeared.

The authors of this study, together with some research assistants, were in charge of administering all the tests. Additionally, we were also in charge of performing the observations, which took place on the day of the pretest in the SA context (for practical reasons concerning trip arrangements) and in the third week of class in the AH context.

2.5. Analyses

For the oral narrative, the CLAN (Computerized Language Analysis) program (MacWhinney, 2000) was used to both transcribe and analyze the data according to the measures described in the previous section. Statistical analyses were performed with the Statistical Package for the Social Sciences (SPSS, 2011) for all the measures included in this study.

Since we have two groups of students who did the same tasks twice, the statistical analysis that is the most appropriate for our data is the Repeated Measures Analysis of Variance (RM ANOVA). The between-subjects variable is *context* (SA/AH) and the within-subject variable is *time* (pretest/posttest scores). We examined both the effect of these two variables and their interaction, which in fact is the analysis that would show whether L2 development was the same or different in the two contexts. Before performing the RM ANOVAs, we made sure the data met all the requirements for this test. In the elicited imitation test, we observed that the distribution was not normal, and, as the sample is small, we preferred to use non-parametric tests to explore the results of this task. We performed a Mann-Whitney *U*-Test for between group comparisons and Wilcoxon Signed Rank Tests for within-group comparisons.

4. Results

4.1. Grammaticality Judgment Test (GJT)

Table 2 presents the descriptive statistics for the results of the GJT for the learners in the two contexts in the pretest and in the posttest.

[Table 2]

As can be observed in Table 2, the GJT was very difficult for the two groups of learners, as the average scores in the pre- and posttest were quite low considering that the maximum possible total score was 14. When looking at the means, it is apparent that the scores of the posttest were always higher than those of the pretest. Additionally, the learners in the SA context always outperformed those in the AH context. The results of the RM ANOVA indicate that there was a significant effect of time [$F(1,109)= 13.69$; ***p*<.001**; partial eta2= .112] and context [$F(1,109)= 13.75$; ***p*<.001**; partial eta2= .112] in favor of the SA learners. The interaction between time*context was approaching significance in favor of the AH learners [$F(1,109)= 3.81$; ***p*=.053**; partial eta2= .034], which indicates that, despite their scores being lower than those of their SA peers, they progressed more from pre- to posttest. The effect size was rather moderate for the time factor, and small for context, which indicates that grammar improvement was not uniform across groups.

4.2. Formulaic Sequences Test (FST)

According to the descriptive statistics included in Table 3, this test was also quite difficult for the learners, but not as difficult as the GJT. When analyzing the scores of the two groups, it can be observed that the SA learners always obtained higher scores than the AH learners, and that the pretest scores were always lower than the posttest scores.

[Table 3]

The results of the RM ANOVA indicate that there was a significant main effect of time, with posttest scores being significantly higher than pretest scores considering the whole population [$F(1,111)= 31.04$; $\mathbf{p<.001}$; partial eta2= .219], and a significant main effect of context [$F(1,111)= 9.352$; $\mathbf{p<.001}$; partial eta2= .078] in favor of the SA setting. The effect size was moderate for the time factor, whereas it was rather small for context. The learners in the two contexts progressed in their knowledge of formulaic sequences in comparable ways, as there is no interaction time*context for this task: $F(1,111)= .564$; $p=.454$; partial eta2= .005.

4.3. Oral narrative task

As explained before, this task was used to assess the oral production skills of a subgroup of learners in the two contexts (25 in the AH context and 16 in the SA context). It was analyzed in terms of fluency (syllables per minute), lexical richness (Guiraud's Index), syntactic complexity (clauses per T-unit), and accuracy (percentage of lexically correct nouns and verbs out of the total nouns and verbs used by a particular learner). Table 4

shows the descriptive statistics for the different measures in the pre- and posttest. In this case, we can observe that the learners obtained similar scores in the pretest. As in the previously analyzed tasks, we can still see the tendency for scores in the posttest to be higher than those of the pretest.

[Table 4]

The results of the RM ANOVA indicate that the effect of time was significant in the case of fluency [$F(1,39)= 37.22$; ***p*<.001**; partial eta2= .488], lexical richness [$F(1,39)= 6.89$; ***p*=.012**; partial eta2= .150], and lexical accuracy [$F(1,39)= 5.73$; ***p*=.022**; partial eta2= .128] indicating that the scores in the posttest were significantly higher than those in the pretest. Effect sizes were rather strong for the first measure and moderate for the other two. For syntactic complexity [$F(1,39)= .023$; $p= .880$; partial eta2= .001] and for morphosyntactic accuracy [$F(1,39)= 1.87$; $p= .179$; partial eta2= .046] there was no significant improvement from pre- to posttest [$F(1,39)= .937$; $p=.339$; partial eta2= .023]. There was no main effect of context; consequently, the performance of the learners in abroad and at home settings was comparable in both testing times: fluency [$F(1,39)= .215$; $p=.645$; partial eta2= .005], lexical richness [$F(1,39)= .838$; $p=.366$; partial eta2= .021], syntactic complexity [$F(1,39)= 1.96$; $p=.169$; partial eta2= .048], lexical accuracy [$F(1,39)= .045$; $p=.833$; partial eta2= .001] and morphosyntactic accuracy [$F(1,39)= .372$; $p= .545$; partial eta2= .009]. Finally, one interaction time*context was found in the oral narrative in the area of lexical richness: $F(1,39)= 4.94$; ***p*=.032**; partial eta2= .112, suggesting that the learners in the SA context improved their lexical diversity in their oral production more than those learners who received instruction at home; however, this

improvement was not quite uniform given the small to moderate effect size. For the rest of the areas of oral production under investigation, there was no interaction time*context: fluency [$F(1,39)= .995$; $p=.325$; partial eta2= .025], syntactic complexity [$F(1,39)= .207$; $p=.652$; partial eta2= .005], lexical accuracy [$F(1,39)= .866$; $p=.358$; partial eta2= .022], morphosyntactic accuracy [$F(1,39)= .356$; $p= .554$; partial eta2= .009].

4.4. Elicited imitation task

Finally, we will present the results of the elicited imitation task, which aimed to examine how successful learners were at repeating sentences originally uttered by a native English speaker. The task was measured considering the number of words that were correctly repeated out of the total number of words (119, which were included in 20 sentences). As in the oral narrative, only a subgroup of students in each context could do this task. According to the data presented in Table 5, it can be seen that the learners in the AH context always obtained slightly higher scores than those in the SA context. Additionally, the scores of the posttest were higher than in the pretest for the two groups, although the mean scores in both testing times were quite high (this test can be considered easy for this population).

[Table 5]

The results of the Mann-Whitney *U* test indicate that there were no significant differences between the two groups in the pretest ($U = 188.50$, $p= .923$) or in the posttest ($U = 176$, $p= .656$). When exploring gains from pre- to posttest for each of the groups separately,

the Wilcoxon Signed Rank Test suggests that the learners in the SA context made significant gains in this task from pre- to posttest (Z= -2.81, $\mathbf{p=.005}$) and so did the learners in the AH context (Z= -4.73, $\mathbf{p < .001}$). As a possible way to analyze whether the gains in one context were more significant than in the other, we calculated Cohen's *d*, which was small (0.33) in the SA context and medium (0.67) in the AH context. Considering these results altogether, then, it can be concluded that there were no significant differences between the two contexts under research in the elicited imitation task.

5. Discussion

The aim of this study was to compare two contexts that offer short-term intensive exposure to L2 learners: a summer stay abroad in the target language country (UK), and an intensive summer course in the students' home country (Spain). Our research questions aimed at investigating whether L2 English learners made significant gains after a short-term intensive course whether abroad or at home, and whether one context was more beneficial than the other for L2 development. In order to do that, we used several tests that examined different L2 areas (grammar, formulaic sequences, fluency, lexical richness, syntactic complexity, accuracy, etc.), and we employed a variety of measures to assess learners' performance. The results of our analyses indicate that the two contexts under examination were highly beneficial for L2 development, as the scores in the posttest in most measures tended to be significantly higher than those in the pretest considering the whole population of our study. With respect to our second research

question, our results suggest that generally speaking, both contexts were equally beneficial, as there were few interactions time*context in the RM ANOVAs. The only interactions that were found refer to grammar as assessed in the GJT, for which the AH context was more beneficial (the difference between the contexts approaching significance), and lexical richness in oral production, for which progress was clearer in the SA context.

The results considering the effect of time (differences between learners' performance before and after an intensive L2 experience) suggest that learners improved their L2 skills significantly in all the areas examined, except for syntactic complexity in oral production. Considering the variety of measures used in the present study to analyze English language skills, we could conclude that both programs led to improved L2 performance, despite the fact that they were quite short. The lack of improvement in syntactic complexity is coherent with what other studies have reported for college students (Pérez-Vidal & Juan-Garau, 2009). The other measure for which no significant progress was captured in any of the programs was verb morphosyntactic accuracy, which refers to errors in verb tense and/or agreement. It seems that, although learners significantly improved their ability to judge as ungrammatical a series of sentences included in the GJT, that grammar knowledge was not reflected in learners' oral production. This result could be due to a trade-off effect, as learners might have been focusing on other aspects in their oral production, such as fluency (or vocabulary) and not on others, such as complexity or morphosyntactic accuracy (Skehan, 1998). Our results concerning teenagers are also in line with other studies that have examined adult learners in that fluency and lexical richness develop significantly after some time abroad (Foster,

2009; Freed, 1995; Trenchs-Parera, 2009) or after an intensive course in the case of lexical richness for intermediate learners (Serrano, 2011).

Considering our findings, we can claim that the two contexts led to significant L2 gains. One reason for the success of the programs could be the fact that they provided intensive input as well as many opportunities for output and interaction on a daily/weekly basis (at least 22.5 hours a week). Contexts that encourage intensive L2 exposure have been claimed to facilitate language development (Muñoz, 2012)[3]. According to Serrano (2011), concentrated and continuous exposure to the L2 can facilitate learning, as learners can more easily retrieve the L2 knowledge they acquire. When there are long time lapses between L2-learning episodes, forgetting is more likely to occur.

Concerning the interactions time*context found (marginally significant in favor of the AH context in the GJT and significant in favor of SA context in lexical richness in oral production), they can be attributed to the type of exposure that the learners received. We observed that in the instructional AH setting there was more focus on forms than in the SA setting, which could have led to the slight advantage of AH learners over SA learners on the GJT. Although the teachers at home used communicative activities, games, videos, etc. to encourage learners to use the language and be entertained, the course was still structured around a textbook and a grammatical syllabus, which had a certain focus on language forms. Nevertheless, it should be emphasized that the differences between the two groups were marginally significant, and that the learners AH did not make significant gains in the other measure related to grammar knowledge, which is verb morphosyntactic accuracy.

On the other hand, the learners abroad did not receive explicit grammar instruction. The classes we observed did not emphasize grammar, but rather vocabulary (both explicitly and through communicative tasks). Apart from this fact, these learners could interact in English more often than their AH counterparts, both in the classroom (they used it with other international students) and outside. These opportunities to interact more often in the L2 that SA learners had (together with the fact that they were exposed to more input outside the class) might have led to a more significant improvement in oral production, more specifically, regarding lexical development. This finding is in line with previous studies on L2 gains in the SA context (Dewey, 2008; Foster, 2009).

Nevertheless, the advantage of the SA learners over the AH learners is negligible, considering the number of measures used in this study. One reason for this result might be that the SA participants did not take full advantage of the increased number of opportunities they had to use English. Although, generally speaking, most students used English in class, they tended to use their L1 during breaks, as they used to spend that time with their L1-speaking peers. This preference was also clearly prominent in the dining room, where almost all the students (not only those included in the present study) sat next to others who spoke their L1. In the afternoon activities and in the residence halls, there was a little bit more mixing between students of different nationalities and thus slightly more English use.

The learners' answers to the self-reported questionnaire also confirmed what we observed. Although these learners reported to use the L2 often/almost always in the English class, they did not use the target language as often in other contexts, especially in the dining room and during their free time and weekends.

To learn more about how often the SA participants interacted in English, we asked them to name the two friends and two adults they interacted with the most during their stay, and then specify the language they used with them. For most learners their closest friends were Spanish; thus, they used their L1 (Catalan and/or Spanish) most of the time they interacted. Only 20% of the participants said that they spoke English with their closest friend, typically another international student. This pattern is only normal, as the short stay abroad did certainly not make it easy for students to make many friends outside their groups. Thus, length of stay abroad is an important variable to consider when examining the amount and type of interaction with native speakers (or other international speakers) that SA learners can be expected to engage in. Regarding interaction with adults, there is a substantial number of students who reported to have had an English-speaker as the adult with whom they spoke the most (in English). Unsurprisingly, this adult was typically a teacher or a group leader in the afternoon activities. However, for almost half of the students (40%), the adult they interacted with the most was Spanish (usually one of the coordinators who accompanied them).

Bearing in mind the information we obtained through our observations and through learners' self-reported questionnaires regarding L2 use abroad, we can conclude that not all learners took advantage of all the opportunities they had to use the L2, and there was individual variation concerning the degree and type of L2 use the SA learners made. These findings regarding teenage learners are in line with what Ranta and Meckelborg (2013) suggest in the case of adults. It must be emphasized, however, that teenagers have less freedom to move outside the networks in which they are supervised than adults, and the opportunities to engage with L2 community members are probably

more restricted for the younger group. Having said that, it should be acknowledged that the particular program under analysis may not have facilitated L2 use as much as other programs where the L2 learners live with families or where an effort is made so that there are no students with the same L1 in the same class (as it was for the child participants in Llanes & Muñoz, 2013 mentioned before). Nevertheless, even in programs that theoretically create more opportunities for L2 use, there are many individual differences that will affect the type of experience (linguistic, social, personal, etc.) that learners have in the SA context (DeKeyser, 2014).

However, despite the fact that the SA participants did not use English as much as they could have, the in-class and out-of-class exposure of the AH learners was still less frequent. AH learners generally used English with their teachers in class but only sometimes with their peers. Out-of-class exposure (including breaks and L2 use at home) was clearly limited and mostly restricted to homework. Considering this information, L2 use could partly explain the results obtained in this study.

6. Conclusion

The findings from our study suggest that intensive exposure to a foreign language leads to significant L2 improvement, no matter how short that period is (under three weeks here). Additionally, our research suggests that, although some grammar aspects might be more quickly developed in a more traditional instructional setting and lexical richness abroad, both settings are comparable in the sense that they both encourage L2 gains on the part of the students. Our findings apply to the programs we have analyzed, which are

typical summer programs for teenagers, but not the only ones that exist. Moreover, it must be emphasized that we are dealing with a younger population than typically analyzed in the literature examining learning context, with more restrictions on interactions and building social networks outside their SA program. We are aware of the limitations of our study, among which we could include the fact that we have only used a few measures of L2 performance, despite having used different types of instruments. Another limitation is the limited time allotted for the data collection; thus, only a subgroup of participants could do the audio-oral tests, which were performed on an individual basis.

Nevertheless, despite the limitations of this study, we have reported how different L2 areas develop after two short intensive programs for teenagers, which are highly common nowadays in Europe to learn English and for which little research exists. These programs usually mean a high economic investment on the part of the parents. It is important to know whether learners make any progress in their L2, which is certainly one of the main reasons children and teenagers enroll in an L2 learning program abroad or at home. Our study is one of the first to compare these two programs, and more research is necessary, including other types of programs that have public funding and are addressed to a wider population of young students. Future studies should also investigate retention, which is something we unfortunately could not examine (and which is something hard to investigate, as students who enroll in summer programs are harder to keep track of than participants attending a school program). It could be that the L2 gains obtained after short-term concentrated exposure are more easily lost (as some studies have suggested especially in the field of cognitive psychology, Bahrick and Hall, 2005). On the other

hand, some studies examining retention of L2 gains after a stay abroad report rather positive results (Juan-Garau, 2014; Llanes, 2012). All in all, more research in this area would be valuable as it would make us better equipped to give advice to parents and program organizers concerning short intensive L2 learning programs.

Endnotes

[1] Although the AS-Unit has often been used to segment oral data as an alternative to the T-unit (Foster, Tonkyn & Wigglesworth, 2000), the data in this study was obtained through a monologic picture description task and thus did not elicit the production of segments that are difficult to analyze considering the T-unit.

[2] For the two accuracy measures, two raters coded 15% of the data, and since inter-rater reliability was high (Cronbach alpha > .80), only one rater coded the remaining narratives.

[3] In order to discard a possible task repetition effect –which could have possibly led to significant gains because of performing the same tasks twice within only 18 days– and to further examine the effect of intensity, we recruited a group of EFL learners (n= 44) with a comparable background to the ones included in the present study to do the two closed written tests (GJT and FST) twice, as they were probably more prone than the open tests to show task repetition effects. The analyses showed that learners receiving a maximum of four hours of English instruction per week during the same period of time (18 days) did not make significant gains in grammar or formulaic sequences, thus discarding a task repetition effect and providing indirect evidence for the positive effect of intensive L2 exposure (Author, 2013).

References

Allen, H., & Herron, C. (2003). A mixed-methodology investigation of the linguistic and affective outcomes of summer study abroad. *Foreign Language Annals, 36,* 370–385.

Author (2013)

Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language, 52,* 566–577.

Bardovi-Harlig, K., & Bastos, M. T. (2011). Proficiency, length of stay, and intensity of interaction and the acquisition of conventional expressions in L2 pragmatics. *Intercultural Pragmatics, 8*, 347-384.

Bowden, H. W. (2007). Proficiency and second-language neurocognition: A study of Spanish as a first and second language. Unpublished PhD thesis, Georgetown University, Washington, DC.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Investigating complexity, accuracy and fluency in SLA* (pp. 21-46). Amsterdam/Philadelphia: John Benjamins.

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

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 Publishing, Ltd.

Collins, L., & White, J. (2011). An intensive look at intensity and language learning. *TESOL Quarterly, 45*, 106-133.

Cubillos, J., Chieffo, L., & Fan, C. (2008). The impact of short-term study abroad programs on L2 listening comprehension skills. *Foreign Language Annals, 41*, 157–185.

DeKeyser, R. M. (2007). Study abroad as foreign language practice. In R. M. DeKeyser (Ed.), *Practice in second language: Perspectives from applied linguistics and cognitive psychology* (pp. 208-226). Cambridge: Cambridge University Press.

DeKeyer, R. M. (2010). Monitoring processes in Spanish as a second language during a study abroad program. *Foreign Language Annals, 43*, 1, 80-92.

DeKeyser, R. M. (2014). Research on language development during study abroad: Methodological considerations and future perspectives. In C. Pérez Vidal (Ed.) *Language acquisition in study abroad and formal instruction contexts* (pp. 313-325). Amsterdam: John Benjamins.

Dewey, D. P. (2008). Japanese Vocabulary Acquisition by Learners in three contexts. *Frontiers: The Interdisciplinary Journal of Study Abroad, XV*, 127-148.

Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language Instruction* (pp. 33-68). Cambridge: Cambridge University Press.

Evans, M., & Fisher, L. (2005). Measuring gains in pupils' foreign language competence as a result of participation in a school exchange visit: The case of Y9 pupils at

three comprehensive schools in the UK. *Language Teaching Research*, *9*, 173–192.

Foster, P. (2009). Lexical diversity and native-like selection: The bonus of studying abroad. In B. Richards, M. Daller, D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition* (pp. 91-106). Hampshire: Palgrave Macmillan.

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

Freed, B. F. (1995). What makes us think that students who study abroad become fluent? In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123-148). Amsterdam/Philadelphia: John Benjamins.

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

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, *26*, 275-301.

Gallimore, R., & Tharp, R. G. (1981). The interpretation of elicited sentence imitation in a standardized context. *Language Learning, 31*, 269-392.

Heaton, J. (1966). *Composition through pictures*. Essex: Longman.

Housen, A. (2012). Time and Amount of L2 Contact Inside and Outside the School ? Insights from the European Schools. In C. Muñoz (Ed.), *Intensive Exposure*

Experiences in Second Language Learning (pp. 111–140). Clevedon: Multilingual Matters.

Housen, A., Schoonjans, E., Janssens, S., Welcomme, A., Schoonheere, E., & Pierrard, M. (2011). Conceptualizing and measuring the impact of contextual factors in instructed SLA. The role of language prominence. *International Review of Applied Linguistics, 49,* 83-112.

Ife, A., Vives, G., & Meara, P. (2000). The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics, 4,* 55-84.

Isabelli-García, C. (2006). Study abroad social networks, motivation and attitudes: Implications for second language acquisition. In M. A. DuFon, & E. Churchill (Eds.), *Language learners in study abroad contexts* (pp. 231-258). Clevedon: Multilingual Matters.

Juan-Garau, M. (2014). Oral accuracy growth after formal instruction and study abroad. Onset level, contact factors and long-term effects. In C. Pérez Vidal (Ed.) *Language acquisition in study abroad and formal instruction contexts* (pp. 87-110). Amsterdam: John Benjamins.

Kinginger, C. (2008). Language learning in study abroad: Case histories of Americans in France. *Modern Language Journal Monograph Series*, 1.

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

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, 3*, 189-215.

Llanes, À., & Muñoz, C. (2013). Age effects in a study abroad context: children and adults studying abroad and at home. *Language Learning, 63*, 63-90.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

McManus, K. Mitchell, R. & Tracy-Ventura, N. (2014). Understanding insertion and integration in a study abroad context: The case of English-speaking sojourners in France. *Revue Française de Linguistique Appliquée, 19*, 97-116.

Mora, J. C., & Valls-Ferrer, M. (2012). Oral Fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly, 46*, 610–641.

Muñoz, C. (Ed.) (2006). *Age and the rate of foreign language learning.* Clevedon: Multilingual Matters.

Muñoz, C. (Ed.) (2012). *Intensive exposure experiences in second language learning*. Clevedon: Multilingual Matters.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Ortega, L. (2000) Understanding syntactic complexity: The measurement of change in the syntax of instructed L2 Spanish learners. Unpublished PhD thesis, University of Hawaii at Manoa.

Polanyi, L. (1995). Language learning and living abroad: Stories from the field. In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 271-292*)*. Amsterdam/Philadelphia: John Benjamins.

Pérez-Vidal, C. & Juan-Garau, M. (2009). The effect of study abroad on written performance. *Eurosla Yearbook, 9*, 269–295.

Ranta, L., & Meckelborg, A. (2013). How much exposure to English do international graduate students really get?: Measuring language use in a naturalistic setting. *The Canadian Modern Language Review, 69*, 1-33.

Rodrigo, V. (2011). Contextos de instrucción y su efecto en la comprensión auditiva y los juicios gramaticales: ¿Son comparables cinco semanas en el extranjero a un semestre en casa*? Hispania, 94*, 502–513.

Spada, N. and Lightbown, P. M. (1989). Intensive ESL programs in Quebec primary schools. *TESL Canada Journal, 7*, 11-32.

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

Sanz, C. (2014). Contributions of Study Abroad Research to Our Understanding of SLA Processes and Outcomes: The SALA Project, an appraisal. In C. Pérez Vidal (Ed.) *Language acquisition in study abroad and formal instruction contexts* (pp. 1-16). Amsterdam: John Benjamins.

Serrano, R. (2011). The time factor in EFL classroom practice. *Language Learning, 61*, 117-143.

Serrano, R., Llanes, À., & Tragant, E. (2011). Analyzing the effect of context of second language learning: Domestic intensive and semi-intensive courses vs. study abroad in Europe. *System, 39*, 2, 133-143.

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

Spada, N., & Lightbown, P. M. (1989). Intensive ESL programs in Quebec primary schools. *TESL Canada Journal, 7*, 11-32.

SPSS 20.0 for Windows [Computer software] (2011). IBM Corporation.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass, & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley, MA: Newbury House.

Taguchi, N. (2013). Production of routines in L2 English: Effect of proficiency and study abroad experience. *System, 41*, 109-121.

Taguchi, N. (2015). "Contextually" speaking: A survey of pragmatic learning abroad, in class, and online. *System, 48*, 3-20.

Tracy-Ventura, N., McManus, K., Norris, J., & Ortega, L. (2014). "Repeat as much as you can": Elicited imitation as a measure of oral proficiency in L2 French. In P. Leclercq, H. Hilton, & A. Edmonds (Eds.), *Proficiency assessment issues in SLA research: Measures and practices* (pp. 143-166). Clevedon: Multilingual Matters.

Trenchs-Parera, M. (2009). Effects of formal instruction and a stay abroad on the acquisition of native-like oral fluency. *Canadian Modern Language Review, 65*, 365-393.

Trofimovich, P., Lightbrown, P. M., Halter, R. H., & Song, H. (2009). The development of L2 pronunciation in a listening and reading program. *Studies in Second Language Acquisition, 31*, 609-639.

van Hout, R., & Vermeer, A. (2007). Comparing measures of lexical richness. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary* (pp. 93-115). Cambridge: Cambridge University Press.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing, 17,* 65-83.

White, J., & Turner, C. E. (2005). Comparing children's oral ability in two ESL programs. *The Canadian Modern Language Review, 61,* 491-517.

Tables

Table 1 Age distribution in the two contexts

	13 years old	14 years old	15 years old	16 years old	17 years old
AH	22	13	18	1	4
SA	9	20	10	13	2

Table 2 Descriptive statistics GJT

Context	Pretest	Posttest
	Mean /14 (SD)	Mean /14 (SD)
AH (N=58)	5.58 (2.16)	6.65 (2.46)
SA (N=53)	7.66 (2.75)	7.99 (3.02)

Table 3 Descriptive statistics FST

Context	Pretest	Posttest
	Mean /16 (SD)	Mean /16 (SD)
AH (N=58)	7.78 (3.33)	8.78 (3.06)
SA (N=54)	9.39 (3.58)	10.71 (3.41)

Table 4 Descriptive statistics oral narrative

	Context	Pretest	Postest
		Mean (SD)	Mean (SD)
Fluency	AH (N=25)	86.11 (26.60)	104.23 (28.30)
(Syllables per minute)	SA (N=16)	85.03 (29.29)	98.11 (29.85)
Lexical richness	AH (N=25)	4.58 (0.63)	4.62 (0.66)
(Guiraud's Index)	SA (N=16)	4.19 (0.73)	4.59 (0.66)
Syntactic complexity	AH (N=25)	1.66 (0.41)	1.68 (0.32)
(Clauses per T-unit)	SA (N=16)	1.57 (0.54)	1.48 (0.34)
Lexical accuracy	AH (N=25)	89.79 (8.29)	91.19 (6.62)
(% Correct nouns and verbs)	SA (N=16)	89.34 (8.20)	92.54 (4.92)
Morphosyntactic accuracy	AH (N=25)	.319 (.228)	.351 (.238)
(Verb tense and agreement errors per T-Unit)	SA (N=16)	.253 (.207)	.335 (.318)

Table 5 Descriptive statistics elicited imitation

Context	Pretest	Posttest
	Mean /119 (SD)	Mean /119 (SD)
AH (N=24)	105.04 (7.34)	109.50 (5.98)
SA (N=16)	103 (12.31)	106.81 (10.69)