The long-term effects of study-abroad experience during childhood on English proficiency

Dr. Kholoud A. Al-Thubaiti

Assistant Professor of Applied Linguistics
Department of English, Umm Al-Qura University

The long-term effects of study-abroad experience during childhood on English proficiency

Dr. Kholoud A. Al-Thubaiti

Abstract:

Although study-abroad experience has been shown to be beneficial in foreign-language learning, we still do not know whether such beneficial gains will last in the long run (Llanes & Muñoz, 2013). This study examines whether prior study-abroad experience in an English-speaking country (ESC) during childhood can show long-term linguistic advantages even after return to the home country. The study was conducted in Saudi Arabia with 48 adult Saudi Arabic EFL speakers. They were divided into two groups based on the context at first exposure: study-abroad in an ESC (n=26), and domestic study in Saudi Arabia (n=22). In addition to the context of exposure, two additional experiential factors were examined: age at first exposure (1-13 years) and years of exposure (8-36 years). The linguistic performance of the groups was compared on three tasks: (a) the Oxford Placement Test as a general proficiency measure, (b) the Vocabulary Levels Test devised by Nation (1990) as a measure of vocabulary size, and (c) the Grammaticality Judgment Test as a measure of morphosyntactic knowledge. The results showed a strong interaction effect between years of exposure and prior study-abroad experience

in predicting the level of general proficiency and vocabulary size, but not in morphosyntactic knowledge.

Keywords: study-abroad experience, instructed setting, years of exposure, age at first exposure, general proficiency, vocabulary size, morphosyntax

Dr. Kholoud A. Al-Thubaiti

تأثير التعلّم في الخارج في الصغر على المدى البعيد على تنمية مهارات اللّغة الإنجليزيّة

د. خلود بنت عبدالله عايض الثبيتي

الملخص:

على الرّغم من وجود دراسات سابقة تشير إلى أنّ هناك فوائد جمّة للتعلّم في الخارج على تنمية المهارات اللّغويّة المكتسبة، فإنّ هذه الدّراسات لم تتناول قياس مدى استمراريّة مثل هذا التّأثير الإيجابيّ على المدى البعيد (يانس وميونث ٢٠١٣). فهدفت الدراسة الحالية إلى قياس تأثير التعلّم في الخارج في الصغر على المدى البعيد على مستوى المهارات اللغوية المكتسبة حتى بعد العودة إلى أرض الوطن. فتكونت عينة الدراسة من ٤٨ سعودياً بالغاً متحدثاً للّغة الإنجليزيّة كلغة أجنبيّة. وقسمت عينة الدراسة إلى مجموعتين على النحو التالي: المجموعة الأولى تعلّمت في الخارج، وكان عددهم ٢٦ متعلّمًا، بينما تعلمت المجموعة الثّانية داخل المملكة العربيّة السّعوديّة، وكان عددهم ٢٢ متعلّمًا. فضلاً عن دراسة متغيّر مكان تعلّم اللّغة الإنجليزيّة (في الخارج مقارنة بالدّاخل) فقد تمّ كذلك قياس تأثير متغيريّ العمر عند بدء تعلّم اللّغة الإنجليزيّة (١–١٣ سنة)، وسنوات التّعرّض للّغة (٨–٣٦ سنة). وقد تمّت مقارنة المهارات اللّغويّة للمجموعتين من خلال أدائهم في ثلاثة اختبارات؛ هي: اختبار أكسفورد لتحديد المستوى اللّغويّ العامّ، واختبار حصيلة المفردات اللّغويّة (نيئشن ١٩٩٠)، واختبار الحكم النّحويّ. وقد أظهرت النّتائج أنّ الأثر الإيجابيّ للتعلّم في الخارج مرتبط بسنوات التّعرّض للّغة؛ فكلّما زادت سنوات التّعرّض تعزّز الأثر الإيجابيّ على المدى البعيد. وكان هذا الأثر واضحًا في اختبار تحديد المستوى اللّغويّ العامّ، وحصيلة المفردات اللّغويّة. ولكنّه لم يظهر في اختبار الحكم النّحويّ.

الكلمات المفتاحيّة:

التعلّم في الخارج، بيئة تعليميّة، سنوات التّعرّض للّغة، العمر عند بدء تعلّم اللّغة، المهارات اللّغويّة العامّة، الحصيلة اللّغويّة من المفردات، النّحو والصّرف.

Introduction:

In age-related research, there is consensus that the context of exposure is a crucial factor to scrutinize because of the different age-related benefits documented in naturalistic and instructed settings (Muñoz, 2008). Research has shown consistently that the age factor works differently depending on the characteristics of the context of exposure, and thus the quantity and quality of second-language (L2) input. Although younger learners have been found to excel in a naturalistic setting (e.g., Johnson & Newport, 1989, 1991; Patkowski, 1980), older learners have been found to excel in a classroom setting (e.g., García-Mayo & Lecumberri, 2003; Muñoz, 2006b; Pfenninger & Singleton, 2017). Muñoz (2006a) was among the first to draw attention to the interaction between age at onset and context of exposure, and thus warned against overgeneralizing the 'younger the better' conclusion drawn from age-related studies in naturalistic settings to instructed ones. In fact, research worldwide has consistently shown no considerable linguistic advantages for early instruction in a school context; for example, Al-Thubaiti (2010, 2014) for Saudi Arabia, Larson-Hall (2008) for Japan, Muñoz (2011) for Catalonia, and Pfenninger (2014) for Switzerland. Researchers have argued that the lack of linguistic advantages in an instructed setting is potentially due to the lack of quality input and limited exposure. DeKeyser (2003) argued that child learners process language implicitly and therefore require

exposure to rich language input which would allow them to successfully acquire the grammatical rules of the foreign language. In the absence of sufficient quality input, child learners are often regarded as 'disadvantaged' and 'slow' L2 learners. On the other hand, teenage learners have been found to be more advantaged in a classroom setting because they are cognitively more mature and can learn the language explicitly. Moreover, Larson-Hall (2008) argued that only with increased exposure to L2 input can potential age effects possibly emerge in an instructed setting.

The aim of this study is to contribute to the current body of research by testing the effect of study-abroad experience as one type of classroom exposure but in a naturalistic context. The term 'study-abroad' is used to mean exposure to formal and natural input of the target L2 inside and outside the classroom. The primary goal of this study is to examine whether prior study-abroad experience in an English-speaking country (ESC) during childhood can show long-term linguistic advantages even after return to the home country. Based on the factor of context at first exposure, the study compares the English proficiency of two adult L2 groups: one group had been first exposed to English through study-abroad experience in an ESC, and the other group had been first exposed to English through a domestic L2 classroom experience in Saudi Arabia. The study-abroad group represents a population of temporary stay-abroad residents who accompanied their parents who

were pursuing their higher education abroad. It is quite interesting to test whether such a temporary stay-abroad experience would bring similar or different results from studies with immigrant populations in naturalistic settings (e.g., Johnson & Newport, 1989, 1991; Patkowski, 1980). The English proficiency of the two L2 groups was assessed by three dependent measures of linguistic knowledge: the Oxford Placement Test (OPT), the Vocabulary Levels Test (VLT), and the Grammaticality Judgment Test (GJT).

The paper is organized as follows. First, age-related studies in different contexts of exposure are reviewed. Then, the aim and the research questions of the present study are presented, followed by the methodology, and the results. Finally, the results are discussed in the light of the research questions.

Literature review:

In age-related studies, most research has been conducted with L2 learners (hereafter 'L2ers') in naturalistic and instructed contexts, with very few studies examining the effect of study-abroad experience (Llanes, 2011). The main findings on the long-term effects of starting age on L2 proficiency will be reviewed from the three contexts of exposure: naturalistic and instructed followed by the context of study-abroad.

The naturalistic context of exposure is often characterized by immersion in the target L2 community, where it is presupposed that L2ers are surrounded by ample amounts of L2 input.

The L2ers therefore have a unique opportunity for exposure to natural language spoken by native speakers. In naturalistic contexts, research has been mainly conducted with populations of immigrants and therefore has examined age on arrival as a predictor of performance (e.g. Johnson & Newport, 1989, 1991; Patkowski, 1980). Research on the ultimate attainment in L2 acquisition has shown that the younger learners outperformed the older ones in the long run. Patkowski (1980) was among the first studies which examined the long-term age effects of initial L2 exposure in a naturalistic context. He tested 67 immigrants to the US with at least five years of residence. Two native speakers were asked to rate the nativelikeness of samples of English interviews with all participants (including the US immigrants and fifteen native controls). Five-minute samples from all of the participants were randomized and transcribed to eliminate pronunciation effects on the raters' judgments. The rating scale ranged from 0 (indicating 'no ability') to 5 (indicating 'native-like performance'). The results showed those who had arrived in the US before fifteen years of age were more likely to be judged as native-like than those who had arrived after the age of fifteen. More evidence showing advantages for younger learners in a naturalistic context emerged from the influential study by Johnson and Newport (1989), who tested 46 Chinese and Korean immigrants with at least three years of residence. They were tested on twelve morphological and syntactic rules of English (such as word order, determiners, movement constraints, number

and tense markings, and pronouns). The results showed a strong relationship between age on arrival and accuracy on a grammaticality judgment task. Those who had arrived in the US before the age of seven performed in a native-like manner, whereas those who had arrived after the age of seven were non-native-like. Their study has been extensively replicated in different ways (Bialystok & Miller, 1999; Birdsong & Molis, 2001; DeKeyser, 2000; Yeni-Komshian, Flege, & Liu, 2000, among others). However, the replicated studies did not reproduce similar findings to those of the original study by Johnson and Newport (1989). DeKeyser (2000), for example, showed that up to the age on arrival of fifteen, it was possible for immigrants to achieve native-like performance. DeKeyser's (2000) findings aligned well with those of Patkowski (1980) rather than those of Johnson and Newport (1989). The offset of the critical period remains controversial because it has varied from one study to another (Muñoz & Singleton, 2011). This controversy bears on theoretical stances regarding the existence of a critical period in L2 acquisition (see for an overview DeKeyser, 2013). Nonetheless, the most influential finding of age-related research in naturalistic contexts is that the 'younger the better' for successful L2 acquisition.

Turning to the classroom context, it is often described as a context of minimal exposure in terms of the quality and quantity of L2 input that is normally received in the classroom. The target L2 is taught and spoken by non-native speakers

who are often described as having limited proficiency. Research in the classroom context has been conducted with instructed L2 learners and thus has examined age at first instruction as a predictor of performance. Most of this research tested instructed L2ers who were still studying at school (e.g., García-Mayo & Lecumberri, 2003; Muñoz, 2006b), and very few studies have tested the long-term effects of early instruction on foreign language proficiency (e.g., Al-Thubaiti, 2010, 2014; Larson-Hall, 2008; Muñoz, 2011). The findings on the long-term effects do not show linguistic advantages for early instruction in a school context. For example, in Japan, Larson-Hall (2008) conducted a study with 200 Japanese college students. They were tested on a grammaticality judgment test for morphosyntactic knowledge of English, and on a phonetic discrimination task on three sounds (ɹ/l/w) known to be problematic for Japanese L2ers of English. The results showed that linguistic advantages in favor of early instruction can be noted only after exposure to intensive amounts of input, and that it can differ according to the linguistic domain. In the phonetic discrimination task, early instruction showed benefits after a range of 1200-2200 hours of input. However, the morphosyntax required more input compared with phonetic knowledge. In the grammaticality judgment task, the benefits of early instruction did not emerge until after a range of 1600-2200 hours of input. It is crucial to note that the late starters outperformed the early starters on the grammaticality judgment task after just 800 hours of

input. Also, years of studying English, rather than starting age, was found to be a predictor of success on the grammaticality judgment task.

Al-Thubaiti (2010) carried out another empirical study to examine the potential long-term effects of early instruction on English proficiency in the context of Saudi Arabia. Al-Thubaiti tested 132 Saudi college students on two production tasks (a cloze test and gap-filling), one comprehension task, and another grammaticality judgment task. With amount of input being statistically controlled, the results revealed no significant differences in performance between the early- and middle-school starters. Al-Thubaiti argued that lack of linguistic advantages by age of L2 instruction can be related to the minimal amount of input which they have had. Unlike the Japanese students, the Saudi students had an average of 896 hours of input inside and outside the classroom, which is far less than the threshold of 1600 hours for morphosyntax found by Larson-Hall (2008).

In a follow-up study, Al-Thubaiti (2014) examined the long-term effects of the amount of L2 input with the same Saudi instructed L2ers and the same tasks. Information about the amount of L2 exposure was collected through a detailed background questionnaire. The L2 input measures included years of English study, hours of school instruction and private lessons, hours of college instruction, hours of studying outside the classroom, hours of L2 contact and use outside

the classroom, and percentage of first language (L1) use in listening and speaking, reading, and writing. These factors, alongside age of L2 instruction, were regressed in a hierarchal multiple regression model. The inter-correlations were controlled in the model. The results showed that the L2 input measures were stronger predictors of task performance than age of L2 instruction. Specifically, the results showed that recent input at college level with an average of 89.66 hours (20-162) and prior exposure at school and private lessons with an average of 722.54 hours (576-2976) were significant predictors of task performance. However, recent input at college was shown to explain more variance (10%-17%) than prior input at school and private lessons (6%-10%). Also, years of study with an average of 12.14 years (7-20) was a significant predictor but it accounted for the least amount of variance in task performance (1%-6%) compared with recent and prior input. It was also found that length of stay-abroad during summer vacations related positively to performance on the cloze test and grammaticality judgment task but not on the comprehension or production of tense marking.

Muñoz (2011) showed that the input factor is more important than age of L2 instruction. Muñoz conducted a study with 162 bilingual Spanish-Catalan college instructed L2ers of English. She analyzed the following set of L2 input measures: years of exposure, overall hours of curricular and extra-curricular instruction, recent hours of curricular instruction at college,

recent hours of curricular and extra-curricular instruction at college, current L2 contact, and hours of length of stay-abroad exposure. They were tested on three measures: a general proficiency test (the Oxford Placement Test), a lexical test, and a phonetic identification test assessing perception of categorical vowel contrast. The results showed consistently that age of L2 instruction did not relate to performance on any of the three tasks. Controlling the effect of age of L2 instruction, the results from partial correlation analyses showed that recent college instruction with an average of 784.7 hours (162-1620) had a positive relationship with general proficiency and lexical knowledge but not with phonetic knowledge. Years of exposure with an average of 13.9 years (10.6-23.4) had a positive relationship with lexical knowledge but not with general proficiency or phonetic knowledge. Also, current L2 contact related positively to global proficiency and phonetic knowledge, whereas length of stay-abroad related positively to general proficiency, lexical knowledge, and phonetic knowledge (Muñoz, 2011).

As shown from classroom research, the experience of stay-abroad emerged as a factor that positively relates to L2 proficiency. Study-abroad is another form of classroom exposure but in a naturalistic context. According to Collentine (2009, p. 218), study-abroad "takes place in countries where the L2 enjoys an important sociological and functional status, entailing a combination of planned curriculum and a host

family." This form of exposure offers several opportunities for using and practicing the target L2 with native speakers in a natural environment outside the classroom. However, the experience of study-abroad is usually limited in terms of length of stay, which is determined by the length of the study program or the purpose of the travel abroad (such as job opportunities, summer schools, scholarship programs, and exchange programs). Study-abroad is hence characterized by temporary residence in the target L2 community. L2ers are presupposed to be exposed to formal and natural L2 inside and outside the classroom. Study-abroad experience therefore stands between classroom and naturalistic exposure. Research on study-abroad has usually been conducted with classroom L2ers who spend the summer or a year abroad in order to advance their L2 proficiency. They have often been compared to their counterparts from a domestic L2 classroom at home (e.g., Llanes & Muñoz, 2009, 2013; Segalowitz & Freed, 2004). Most of the available literature examined the effect of study-abroad on oral fluency (e.g., Llanes & Muñoz, 2009; Segalowitz & Freed, 2004) and fewer studies tested its effect on vocabulary and grammar (e.g., Collentine, 2004; Llanes, 2010) and global L2 proficiency (e.g., Segalowitz *et al.*, 2004). Most of these studies compared the effect of study-abroad with domestic classroom exposure at home, and sometimes with immersion classroom types. According to Llanes (2011), the overall findings support beneficial gains from study-abroad compared with the domestic classroom context. Nonetheless,

not all aspects of L2 proficiency were examined thoroughly. So far, there is conclusive evidence for positive gains in oral fluency and increased vocabulary growth compared with grammar and morphology. For example, Howard (2006) studied the L2 acquisition of French by Irish college students and found that their morphological development of plural marking was enhanced after study-abroad experience. On the other hand, DeKeyser (1991) showed no substantial differences in grammatical performance between L2ers of Spanish with study-abroad and with domestic classroom experience.

Another crucial factor to consider in this research area is whether any beneficial gains from study-abroad experience lead to short- or long-term effects. So far, there has been limited research on this issue. In fact, Llanes and Muñoz (2013, p. 83) remarked that "additional research is needed regarding the long-term effects of the gains that emerge as a result of an SA [study-abroad] experience, because the duration of these effects is unclear". From a practical perspective, the long-term effects of study-abroad are important to assess because they will inform policy makers and parents of the actual linguistic gains from (financial and time) investment in the study-abroad programs.

To summarize, age-related research has shown that the effect of starting age on L2 proficiency in the long run varies depending on the context of exposure. While there is ample evidence that immersion in a naturalistic setting during

childhood has long-term linguistic advantages, there is little research on the long-term effects of study-abroad experience as a form of classroom exposure but in a naturalistic setting. The aim of the present study is to address the beneficial gains from study-abroad experience in terms of its long-term effects with a different type of L2 speakers, as will be explained in the following section.

The aim and research questions of the study:

This study aims at testing whether prior study-abroad experience in an English-speaking country (ESC) during childhood can show long-term linguistic advantages even after return to the home country. To address this aim, the study compares the English proficiency of adult L2 speakers who had study-abroad experience in an ESC with another group who had domestic L2 classroom instruction in Saudi Arabia. In addition to the country of first L2 exposure (ESC as opposed to Saudi Arabia), the study examines the effects of age at first L2 instruction and length of L2 exposure as interacting factors. Three dependent measures of linguistic knowledge were examined: (a) general English proficiency as measured by the Oxford Placement Test (OPT), (b) breadth of vocabulary size as measured by the Vocabulary Levels Test (VLT), and (c) morphosyntactic knowledge as measured by a Grammaticality Judgment Test (GJT).

Considering the three experiential factors (country of first exposure, years of exposure, and age at first exposure) as potential interacting predictors of L2 proficiency, the research questions of the study are formulated as follows:

1. Which of the L2ers' experiential characteristics could explain their task performance when L2 proficiency is examined by an OPT, VLT, and GJT?

2. In a GJT, which of the three experiential factors could most explain the L2ers' ability to distinguish between grammatical and ungrammatical morphosyntactic conditions?

It is crucial to note that the study-abroad group of this study differs from the type of study-abroad groups normally tested in the current literature in two respects as follows. First, the L2 speakers of the study-abroad group were not enrolled in study-abroad programs but rather in public schools in the native community; second, they had not intended to travel abroad for language purposes, but by accompanying their parents they had the opportunity to acquire English as a second language and interact with local children of their own age. They thus had exposure to the L2 at home, at school, and in the local community. Therefore, the L2 speakers of the study-abroad group bore more similarities to the immigrant populations whose children usually get immersed in the local community through the school system and everyday interaction with the community. However, they differed from immigrants'

children in that they did not have plans for permanent residence and integration in the target community, and that they came from a higher socio-economic status.

Method:

Participants:

The study sample consisted of 48 adult L2 speakers of English. They were highly educated and active bilinguals in L1 Saudi Arabic and L2 English. They were either working or studying at college level in Saudi Arabia. At the time of testing, most of them were majoring in English literature or TESOL with a few in computer science, information science, and mathematics. They represented the same socio-economic group. Their parents were highly educated and most of them spoke English in their professions. The participants were divided into two groups by the country of first exposure: (a) study-abroad in an ESC and (b) domestic study in Saudi Arabia (SA). For reasons of brevity throughout the paper, the first group will be referred to as ESC, and the second group as SA. Table 1 presents a summary of the experiential variables for the two groups.

Table 1. Summary of the experiential variables of the L2 groups (*N*=48)

Groups by country of first exposure	*n*	Age at testing	Age at first exposure	Years of exposure
		Mean (SD) Min-Max	*Mean (SD) Min-Max*	*Mean (SD) Min-Max*
Study-abroad in an ESC	26	26.35 (5.56) 19-37 years.	4.62 (2.61) 1-12 years.	21.73 (6.88) 8-36 years.
Domestic study in SA	22	25.23 (4.15) 20-32 years.	9.73 (3.17) 4-13 years.	15.50 (5.14) 8-26 years.

Research instruments:

Three dependent measures of linguistic knowledge were used: (a) the Oxford Placement Test as a general proficiency measure, (b) the Vocabulary Levels Test devised by Nation (1990) as a measure of vocabulary size, and (c) the Grammaticality Judgment Test as a measure of morphosyntactic knowledge. A description of each test of the three is given below.

Oxford Placement Test (OPT):

The OPT (2001) is a test of English language L2 proficiency in reading, vocabulary, and grammar. The questions were constructed in the multiple-choice format. The test had 60 questions divided in two parts: the first part comprised 40 questions and was designed to be taken by respondents with lower levels of proficiency, and the second part had 20

questions designed for higher levels of proficiency. The questions in part 2 were incrementally harder than those in part 1. Therefore, if participants scored 36 or more in part 1, it was recommended that they should complete part 2 of the test. The test came in two versions: a computer-based test and a paper-and-pen test. The test used in the current study was the paper-and-pen version, and both parts of the test were administered. The test took an average of 30 minutes to complete. The maximum score in the OPT was 60. Examples of one type of question from part 1 and another from part 2 are given in the Appendix.

Vocabulary Levels Test (VLT)[1]:

The VLT (Nation, 1990, 2001) is a measure of the L2ers' vocabulary size of general and academic English. The test version used consisted of five blocks, each representing one of the vocabulary levels (2000 words, 3000, 5000, 10,000, and academic level). It is assumed that each vocabulary level correlates with the ability to use English in different tasks (such as reading a newspaper column and writing an argumentative essay) (Schmitt, Schmitt, & Clapham, 2001). For example, L2ers with knowledge of the 10,000 most frequent words in English are often described as having a 'wide vocabulary' (Schmitt *et al.*, 2001, p. 56) which allows them to cope with advanced uses of English. In the test, each level had six vocabulary sets, with each set consisting of six vocabulary items and three meanings

---

[1] I am grateful to Dr. Suhad Sonbul for providing references on the VLT and answering my questions about vocabulary testing in an effective and prompt manner.

(that is, three test items ✗ six options = eighteen test items per level). The questions were constructed in the form-recognition format (Schmitt, 2010). The participants' task was to match each of the three meanings with their corresponding vocabulary item from the set. The vocabulary items increased in difficulty as the frequency level increased. The maximum score of the VLT was 90. The test version administered was paper-and-pen, and it took an average of 35 minutes to complete. Here is an example of one set which appeared in the instructions section:

1. business

2. clock   _____ part of a house

3. horse   _____ animal with four legs

4. pencil   _____ something used for writing

5. shoe

6. wall

Grammaticality Judgment Test (GJT):

A shortened version of the GJT designed by Al-Thubaiti (2010) was used. The purpose of using a GJT was to assess L2ers' morphosyntactic knowledge. The test consisted of 144 test items and twelve ungrammatical distracters. The overall test material was balanced in grammaticality with half grammatical and half ungrammatical. The ungrammatical items were included to examine violations of specific grammar-constraints in English. If the L2ers could distinguish between grammatical

and ungrammatical conditions, that was evidence of acquiring that specific morphosyntactic constraint (White, 2003).

The test material was administered on a computer under time pressure in aural and written format. On a paper answer sheet, the participants had to rate the test items on a five-point Likert scale ranging from (1) 'definitely impossible' to (5) 'definitely possible'. They were instructed to rate the test items based on their first impression in order to avoid metalinguistic analysis. They had eight seconds to mark their ratings on the answer sheet. Examples of the test items for grammaticality are given in the Appendix.

Background questionnaire:

A background questionnaire was designed to gather information about the L2ers' learning experience of English in terms of their age at testing, age at first exposure, country of first exposure, and years of exposure. Age at first exposure was counted from the time they reported a significant encounter with the English language (Muñoz, 2011). Years of exposure was verified by subtracting age at first exposure from age at testing (Nishikawa, 2014).

Procedure of testing and data analysis:

According to the code of ethics of data gathering, an informed consent was obtained from the participants. The testing took place individually in two sessions in a quiet office. In the first

session, the participants took the GJT because it was computerized and they filled out the background questionnaire. Before starting, the participants had instructions and practice on how to conduct the GJT. In the second session, they undertook the OPT and VLT, for which they also had instructions on how to conduct these tests.

Using the R version (R Core Team, 2018), two statistical analyses were conducted on the L2 data: multiple linear regression and mixed-effects modelling. According to Levshina (2015), multiple regression has the advantage of estimating the effect of each predictor individually while controlling for any potential confounding effects among the predictors, such as age at exposure and years of exposure. The lm () function was used to perform multiple regression, and the visreg package (Breheny & Burchett, 2017) was used to visualize the interactions in the regression model (Levshina, 2015). On the other hand, mixed-effects modelling was chosen for the many advantages it provides (Cunnings & Finlayson, 2015). The most important advantage is that mixed-effects can account for random variance due to variation across different items and participants in one analysis. Another advantage is that it allows modelling continuous and categorical predictors in one model. It can also be performed on the raw data without it being averaged, and thus is not affected by missing data points (For more information see Cunnings & Finlayson, 2015). To conduct linear mixed-effects modelling, the lmer ()

function was used from lme4 package (Bates, Mächler, Bolker, & Walker, 2015).

Results:

The effect of experiential factors on L2 task performance:

Table 2. L2ers' scores by country of first exposure on the three linguistic tasks

	ESC			SA		
	n	*Mean (SD)*	*Min-Max*	*n*	*Mean (SD)*	*Min-Max*
OPT	26	49 (8.24)	35-60	22	50.18 (3.95)	43-59
VLT	26	71.88 (16.75)	31-90	22	77.18 (9.09)	54-88
GJT	26	101.77 (18.82)	68-131	22	103 (13.64)	74-123

Note. ESC= Study-abroad in an ESC; SA= Domestic study in Saudi Arabia; OPT= Oxford Placement Test (maximum score 60); VLT= Vocabulary Levels Test (maximum score 90); GJT=Grammaticality Judgment Test (maximum score 144).

Table 2 shows the L2ers' scores by country of first exposure on the three linguistic tasks: OPT, VLT, and GJT. For each linguistic task, a multiple regression linear model with seven parameters was constructed for testing the L2ers' scores as a function of the three experiential factors: country of first exposure (CoE), age on first exposure (AoE), and years of exposure (YoE). Three two-way interactions were also constructed in the design: the first aimed at testing whether the effect of AoE varied by CoE (CoE: AoE), the second aimed at

testing whether the effect of YoE varied by CoE (CoE: YoE), and the third aimed at testing whether the effect of AoE varied by YoE (AoE: YoE). The categorical predictor of CoE was sum coded as (-.5=ESC,.5=SA) to obtain ANOVA main effects style (Cunnings & Finlayson, 2015). The continuous predictors of AoE and YoE were centered to minimize the collinearity between main effects and interactions in the model (Baayen, 2008). The models and their results are presented next in the following order: Table 3 for the OPT, Table 4 for the VLT, and Table 5 for the GJT.

Table 3. Summary of the predictors' coefficients in a multiple regression model for the OPT

	Estimate	Standard Error	Standardized Beta	*t* value	*p*
(Intercept)	49.44	1.13	0.00	43.92	<.000
s_CoE	1.42	2.09	0.11	0.68	0.501
cAoE	0.57	0.35	0.33	1.62	0.112
cYoE	0.54	0.17	0.56	3.21	0.003
s_CoE:cAoE	-0.46	0.69	-0.10	-0.67	0.507
s_CoE:cYoE	-1.35	0.43	-0.62	-3.10	0.003
cAoE:cYoE	0.07	0.04	0.28	1.70	0.097

Note. Model: lm (formula = OPT ~ s_CoE + cAoE + cYoE + s_CoE:cAoE + s_CoE:cYoE + cAoE:cYoE, data = L2data_w). The categorical predictor s_CoE was sum coded as (-.5=ESC,.5=SA), and the continuous predictors of cAoE and cYoE were centered. Shaded rows indicate a significant predictor or interaction.

For the OPT, the model was significant, $F(6, 41) = 5.854$, $p < 0.0001$. As indicated by the multiple R^2 (0.461), the model accounted for 46.1% of variance in the dataset. This value is considered high. Looking at the fixed effects individually and the interactions, Table 3 shows that not all the coefficients in the model were significant. Of the three fixed predictors, YoE stood out as the significant main predictor of the L2ers' performance in the OPT. However, the main effect of YoE was qualified by a significant interaction with CoE. This is clearly shown in Figure 1. This interaction suggested that the effect of YoE was not the same for the ESC and SA groups. Unlike the SA group, the L2ers' scores of the ESC group increased with increased YoE. It is crucial to note that the effect of AoE did not vary by CoE or by YoE (*see* Figure 1).

(A) CoE:YoE (*p=0.003*)
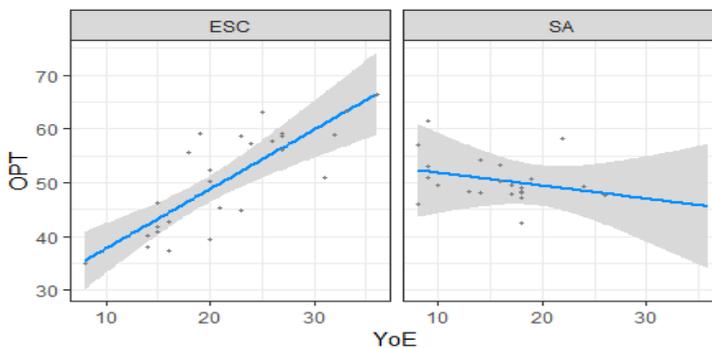
(B) CoE:AoE (p=0.507, non-sig)
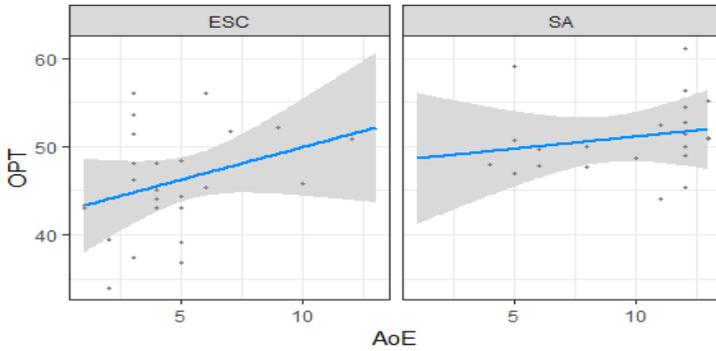
(C) AoE:YoE (p=0.097, non-sig)
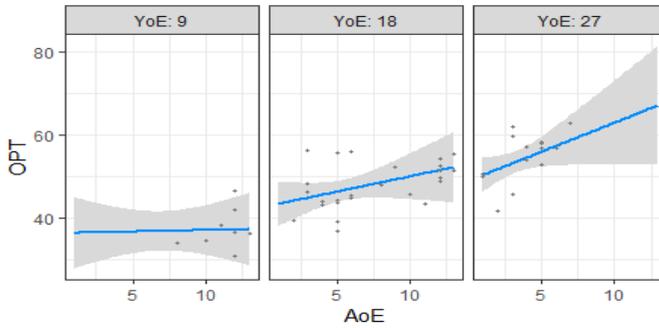
Figure 1. Interaction plots for scores on the OPT: (A) CoE:YoE, (B) CoE:AoE, and (C) AoE:YoE

Table 4. Summary of the predictors' coefficients in a multiple regression model for the VLT

	Estimate	Standard Error	Standardized Beta	t value	p
(Intercept)	76.25	2.42	0.00	31.50	<0.000
s_CoE	7.82	4.49	0.28	1.74	0.089
cAoE	0.79	0.76	0.22	1.04	0.304
cYoE	1.10	0.36	0.54	3.03	0.004
s_CoE:cAoE	-1.45	1.48	-0.15	-0.98	0.334
s_CoE:cYoE	-2.79	0.93	-0.61	-2.99	0.005
cAoE:cYoE	0.22	0.09	0.40	2.40	0.021

Note. Model: lm (formula = VLT ~ s_CoE + cAoE + cYoE + s_CoE: cAoE + s_CoE:cYoE + cAoE:cYoE, data= L2data_w). The categorical predictor s_CoE was sum coded as (-.5=ESC,.5=SA), and the continuous predictors of cAoE and cYoE were centered. Shaded rows indicate a significant predictor or interaction.

Turning to the VLT, the model was also significant, F (6, 41) = 5.375, p <0.0001. As indicated by the multiple R^2 (0.440), the model accounted for 44.0% of variance in L2ers' performance. In terms of the effects of the predictors, Table 4 shows YoE as the only significant main predictor of the L2ers' scores on the VLT. However, this main effect was qualified by two interactions: one with CoE and the other with AoE. As shown in Figure 2, the L2ers in the ESC group, unlike the SA group, gave higher scores with increased YoE. This interaction suggested that the effect of YoE varied by CoE, and that the ESC group had an advantage over the SA group. The results

also showed that there was a significant interaction effect between YoE and AoE. As clearly seen in Figure 2, the interaction here suggested that younger L2ers with more YoE scored higher than younger ones with fewer YoE. On the other hand, AoE did not show a main effect or an interaction effect with CoE (*see* Figure 2).

(A) CoE:YoE (*p*=0.005)
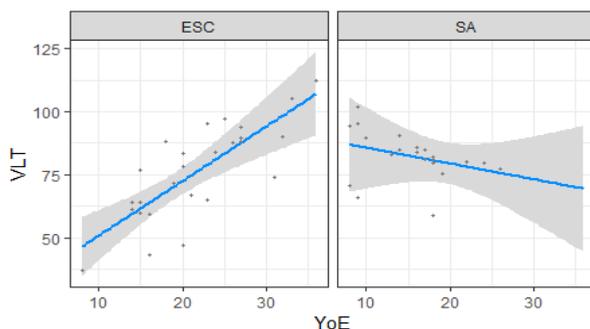
(B) CoE:AoE (*p*=0.334, non-sig)
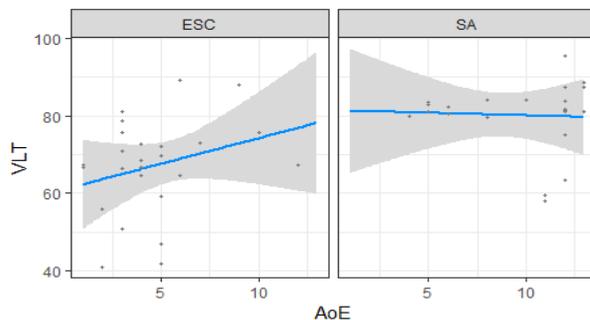
C) AoE:YoE (*p*=0.021, non-sig)
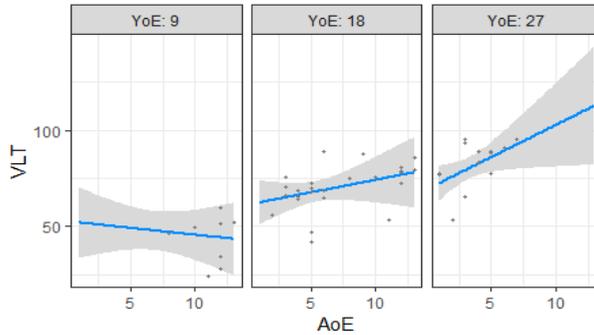
**Figure 2.** Interaction plots for scores on the VLT: (A) CoE:YoE, (B) CoE:AoE, and (C) AoE:YoE

Table 5. Summary of the predictors' coefficients in a multiple regression model for the GJT

	Estimate	Standard Error	Standardized Beta	t value	P
(Intercept)	102.37	3.49	0.00	29.31	<.000
s_CoE	-1.65	6.47	-0.05	-0.26	0.799
cAoE	1.76	1.09	0.41	1.61	0.116
cYoE	1.03	0.52	0.43	1.97	0.056
s_CoE:cAoE	-0.25	2.14	-0.02	-0.12	0.907
s_CoE:cYoE	-1.73	1.35	-0.32	-1.29	0.204
cAoE:cYoE	0.14	0.14	0.21	1.01	0.318

Note. Model: lm (formula = GJT ~ s_CoE + cAoE + cYoE + s_CoE:cAoE + s_CoE:cYoE +cAoE:cYoE, data = L2data_w). The categorical predictor s_CoE was sum coded as (-.5=ESC,.5=SA), and the continuous predictors of cAoE and cYoE were centered.

On the other hand, the GJT revealed different results from the OPT and VLT. The model was non-significant, $F(6, 41) = 1.415, p = 0.232$. As indicated by the multiple R^2 (0.172), the model managed to account for at most 17.2% of the variance in the dataset. The model therefore suggested that the three experiential factors altogether did not contribute substantially to the L2ers' performance. Table 5 shows no significant main effects of AoE or CoE, but a (marginally) non-significant main effect for YoE. There was also no significant interaction effect between the tested predictors (*see* Figure 3). Again, of all the predictors, it was YoE which showed a tendency effect but it was not significant ($p=.056$).

(A) CoE:YoE ($p=0.204$, non-sig)
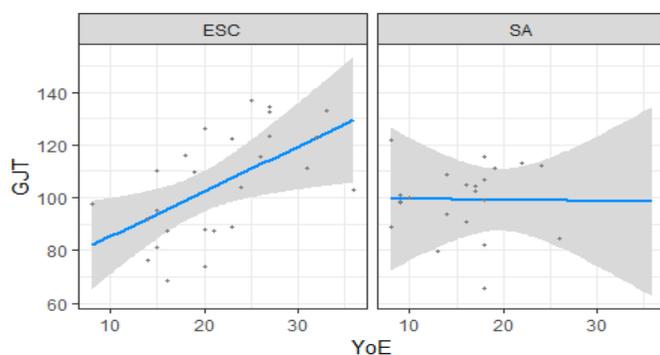
(B) CoE:AoE ($p=0.907$, non-sig)
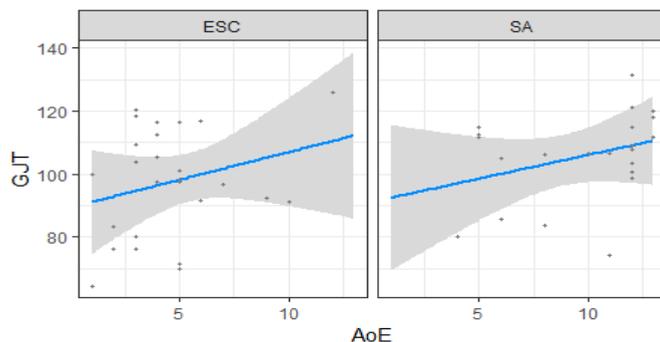
(C) AoE:YoE (*p*=0.318, non-sig)
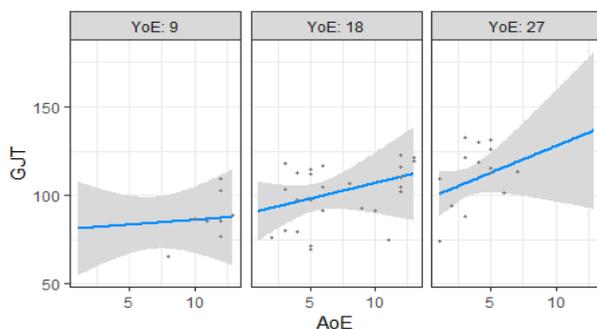
Figure 3. Interaction plots for scores on the GJT: (A) CoE:YoE, (B) CoE:AoE, and (C) AoE:YoE

To summarize, a model of multiple regression with three main predictors (CoE, AoE, and YoE) and three interactions (CoE:AoE, CoE:YoE, and AoE:YoE) can account for 46.1% of variance at most on the OPT, followed by 44.0% on the VLT, and the least of 17.2% on the GJT. Unlike the OPT and the VLT, the L2ers' scores on a morphosyntactic measure such as the GJT was minimally affected by the regressed model. Of the three experiential predictors, YoE stood out as the most predicting factor of L2ers' performance, especially among the ESC group on the OPT and the VLT. The interaction between YoE and CoE was significantly attested in the L2ers' performance on the OPT and the VLT. The predictor of AoE showed no significant effect on any of the three linguistics tasks. Unexpectedly, there was no significant interaction effect between AoE with CoE on the three measures. There was an

interaction effect between YoE and AoE on the VLT measure but not on either the OPT or the GJT.

The effect of experiential factors on the L2ers' ratings of (un)grammaticality in a GJT:

Table 6. Mean ratings (Z-ratings) on two grammaticality conditions by country of first exposure:

	ESC			SA		
	n	*Mean (SD)*	*Z-mean (SD)*	*n*	*Mean (SD)*	*Z-mean (SD)*
Grammatical	2028	4.42 (1.07)	0.5 (0.68)	1716	4.29 (1.15)	0.42 (0.73)
Ungrammatical	1716	2.86 (1.65)	-0.48 (1.04)	1452	2.65 (1.58)	-0.62 (1.00)

Note. ESC= Study-abroad in an ESC; SA= Domestic study in Saudi Arabia; *n*= observations of L2ers in long-data format

Table 6 shows that L2ers from both groups gave higher mean ratings to the grammatical sentences compared with the ungrammatical ones. So, I tested whether their ability to distinguish grammaticality depended on their characteristics in terms of CoE (ESC vs SA), AoE, or YoE. A linear mixed-effects model was fitted using the technique of restricted maximum likelihood (REML). The model was performed with 6912 observations, 144 items, and 48 participants. As recommended by Schütze and Sprouse (2013), the dependent variable of mean ratings was z-transformed for the grammatical and ungrammatical items to normalize the data ratings. Fixed effects included the effects of the 'grammaticality'

condition (g=grammatical vs ug=ungrammatical) as a within-participant factor (that is, a repeated measure), and L2ers' characteristics (AoE, CoE, and YoE) as between-participant factors. The categorical fixed predictors of 'condition' and 'CoE' were sum coded to obtain ANOVA main effects style (Cunnings & Finlayson, 2015). The CoE was sum coded as (-.5=ESC,.5=SA) and the condition was sum coded as (-.5=g,.5=ug). On the other hand, the continuous fixed predictors of AoE and YoE were centered to minimize the collinearity between main effects and interactions in the model (Baayen, 2008). Four two-way interactions were built in the model as follows: three interactions were constructed to measure how the mean ratings by condition interact with each of AoE, YoE, and CoE (s_condition:cAoE, s_condition:cYoE, and s_condition:s_CoE, respectively), and the fourth interaction aimed at testing how variation in AoE is affected by YoE (cAoE: cYoE). Random effects were fit using a 'maximal' random effects structure. This included random intercepts for L2ers and items to model how the overall z-ratings for each L2er and item varied randomly. Random slopes for the fixed repeated measures effect of 'condition' were modelled to vary by both L2ers and items. Statistical significance was assessed by calculating the p values from the t distribution using the following equation: [2 * (1-pt (abs(x), Y-Z))] (Baayen, 2008, p. 248).

Table 7. Mixed-effects model for L2ers' z-ratings on (un)grammatical conditions in the GJT

Parameters	Fixed effects				Random effects by	
					Items	L2ers
	Estimate	Standard Error	*t* value	*p*	SD	SD
Intercept	-0.06	0.06	-1.082	0.279	0.26	0.30
s_condition	-1.01	0.097	-10.329	<0.001	0.52	0.51
s_CoE	-0.20	0.12	-1.652	0.099	-	-
cAoE	0.00	0.02	-0.122	0.902	-	-
cYoE	-0.02	0.01	-1.701	0.089	-	-
cAoE: cYoE	-0.00	0.00	-0.574	0.566		
s_condition: cYoE	-0.05	0.02	-2.942	0.003	-	-
s_condition: s_CoE	-0.09	0.20	-0.451	0.652	-	-
s_condition: cAoE	-0.05	0.03	-1.431	0.152	-	-

Note. Model Formula: z-ratings ~ s_condition + s_CoE + cAoE + cYoE + cAoE:cYoE+ s_condition:cYoE + s_condition:s_CoE + s_condition:cAoE + (1 + s_condition|L2ers) + (1 + s_condition|Items). The categorical fixed factors of condition and CoE were sum coded, as follows: s_condition (-.5=g,.5=ug) and s_CoE (-.5=ESC,.5=SA). The continuous fixed factors of cAoE and cYoE were centered. Shaded rows indicate significant predictors or interactions.

The results showed that the variance of the random effects (intercepts and slopes) altogether accounted for 59.76% of the model, leaving a residual of 40.24 % as unexplained error. Specifically, the results

indicated that among the random effects, the intercepts for L2ers and items accounted for 8.04% and 5.75% respectively, whereas the variance of the L2ers' slope for condition explained 22.72% of the model, and the items' slope for condition explained 23.26% of the model.

As shown in Table 7, the model revealed a significant main effect of condition with the negative estimate confirming that the ungrammatical sentences were rated as significantly less acceptable than the grammatical sentences. The main effect of condition was qualified by a significant interaction effect with YoE. The interaction indicated that the ratings on grammaticality condition differed by YoE. The L2ers with longer periods of exposure gave lower ratings to ungrammatical conditions, but higher ratings to grammatical conditions (*see* Figure 4). Also, the model showed that ratings on grammaticality condition were not affected by an interaction with either CoE or AoE (*see* Figure 5).
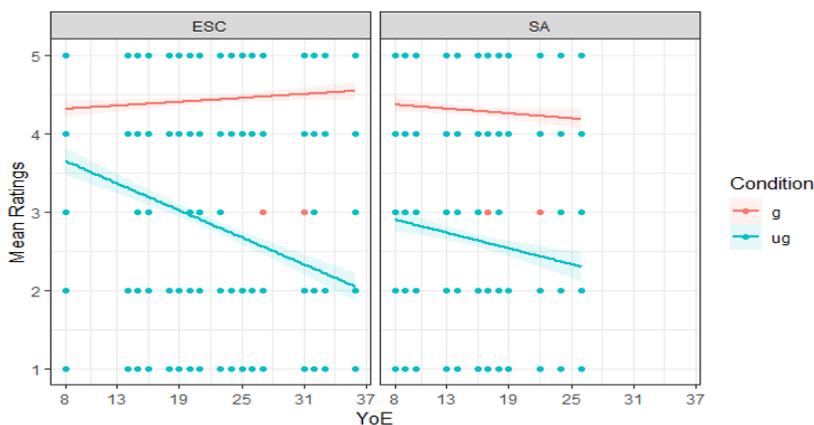

Figure 4. Interaction plots for the mean ratings on the GJT by condition (grammatical vs ungrammatical) depicting the interaction between CoE and YoE
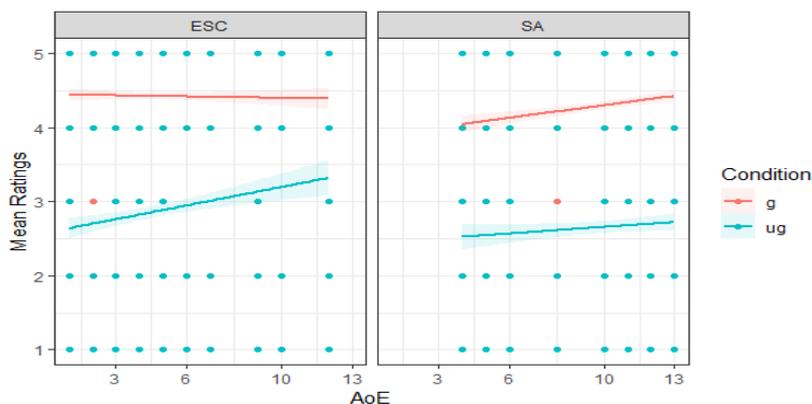
Figure 5. Interaction plots for the mean ratings on the GJT by condition (grammatical vs ungrammatical) depicting the interaction between CoE and AoE

To summarize, mixed-effects modelling with three main fixed predictors (CoE, YoE, and AoE) and four interaction effects (condition:AoE, condition:YoE, condition:CoE, and AoE:YoE) showed that the L2ers' morphosyntactic knowledge was most affected by condition (grammatical vs ungrammatical) interacting with YoE. The L2ers with increased YoE distinguished better between grammatical and ungrammatical conditions. They gave significantly higher ratings to grammatical conditions and lower ratings to ungrammatical ones. On the other hand, the results showed no significant effect for CoE or AoE either as main predictors or as interacting ones with condition.

Discussion:

The primary goal of this study was to examine whether study-abroad experience in an ESC during childhood will show long-term linguistic advantages. To accomplish this goal, two adult L2 groups were compared: those who had study-abroad experience in an ESC and those who only had domestic L2 classroom experience in Saudi Arabia. The groups were referred to as the ESC and SA groups, respectively. Their L2 linguistic knowledge was compared on three measures: (a) the Oxford Placement Test (OPT) as a general proficiency measure, (b) the Vocabulary Levels Test (VLT) as a measure of vocabulary size, and (c) the Grammaticality Judgment Test (GJT) as a measure of morphosyntactic knowledge. Their L2 performance was assessed statistically against three experiential factors, country of first exposure (ESC vs SA), age on first exposure, and years of exposure. Two research questions were formulated: the first was designed to test which of the experiential factors could best predict task performance on the three linguistic measures and the second to assess which of the experiential factors could best predict the L2ers' ability to distinguish (un)grammaticality in a morphosyntactic test.

To answer the first question, multiple regression analyses were performed. The results showed that studying abroad during childhood in an ESC did not alone grant long-term linguistic advantages on any of the three tasks (OPT, VLT, or GJT). Rather, years of exposure was a crucial interacting

factor with study-abroad in predicting general proficiency level and vocabulary size, but not morphosyntax. In the OPT and VLT measures, L2ers from the study-abroad group scored significantly higher with more years of exposure than those with fewer years of exposure. This effect did not emerge among the L2ers with domestic classroom experience in SA. It is interesting to note that the interaction between prior study-abroad and years of exposure was not attested across all three measures. These results concur with those of previous studies on the effect of (current) study-abroad experience, which showed advantages in favor of vocabulary growth (e.g., Collentine, 2004; Llanes, 2010) and general L2 proficiency (e.g., Segalowitz *et al.*, 2004). Muñoz (2011) also found that length of stay-abroad benefitted general proficiency and vocabulary. The results on morphosyntax in previous studies are already mixed. The present findings concur more with those of DeKeyser (1991), who did not find substantial gains in the area of grammar.

To answer the second question relating to morphosyntax, mixed-effects modelling was performed with the fixed factors of country of first exposure, age on first exposure and years of exposure, and the random factors including items and L2ers. The results showed that studying abroad during childhood did not grant L2ers any long-term benefits with their morphosyntactic knowledge; the study-abroad and domestic study groups performed statistically the same. Rather, years of exposure

emerged as the best predictor of L2 task performance. Years of exposure interacted significantly with the L2ers' ability to distinguish grammatical and ungrammatical conditions. As years of exposure increased, L2ers managed to show a sharper distinction between grammatical and ungrammatical sentences. This result suggests that the context of first exposure does not make a difference in L2ers' morphosyntactic gain.

Turning now to discuss the contribution of age at first exposure in a study-abroad context compared with a domestic L2 classroom, and the extent to which age at first exposure could potentially predict L2 performance in the long run. From the set of experiential factors assessed in this study, age at first exposure did not show any significant effect either as a main factor or even as an interacting factor with country of first exposure (ESC vs SA). This finding is unexpected given that a large number of age-related studies have shown that age at first exposure interacts with the context of exposure. Previous studies have shown that 'the younger the better' holds good in a naturalistic setting (e.g., Johnson & Newport, 1989, 1991; Patkowski, 1980), whereas 'the older the better' has been found to be true in an instructed setting (e.g., García-Mayo & Lecumberri, 2003; Muñoz, 2006b; Pfenninger & Singleton, 2017). It was therefore concluded that age is an important factor in a naturalistic setting but not in an instructed setting. However, the current study did not manage to provide any supporting evidence for such an interaction with study-

abroad experience. One possible explanation for this is that the study-abroad group forgot English and therefore did not show any advantages compared with the domestic study group in SA. Although this is a possibility, the fact should not be overlooked that the L2ers of the study sample were active bilinguals in L1 and L2. Therefore, the possibility to have forgotten English is not very convincing. According to their performance in the OPT, their proficiency in English fell in the range of intermediate to high. Before any conclusions can be drawn, future research should compare the short-term and long-term effects of study-abroad experience. This comparison is crucial for validating whether L2ers could have forgotten English after they had returned to their home country. Another possible explanation for the lack of age effects among the study-abroad group might be related to whether the context of exposure was current or prior. Hence, any beneficial gains from prior exposure could be lost in the long run. At the time of testing, the study-abroad group were in a foreign language setting. Therefore, the most recent input which they had was not naturalistic and intense compared with what would be expected in an ESC. The distinction between current and prior exposure here echoes a similar asymmetry in the effect of recent and prior hours of exposure found in previous research. It was found that recent hours of exposure at college level had more effect than prior hours of classroom exposure (Al-Thubaiti, 2014; Muñoz, 2011). If this explanation holds true, then study-abroad during childhood does not produce

long-term effects. It could also be argued that study-abroad is possibly less effective than living in a naturalistic setting. Although study-abroad experience offers an opportunity of exposure to natural language inside and outside the classroom, it still does not offer the same level of exposure as living abroad, as in the situation of immigrants. This difference requires an investigation of the L2ers' affective factors, such as attitude and motivation. These could be potential contributing factors which the current study did not consider.

Overall, the results have shown a strong interaction effect between years of exposure and prior study-abroad experience in predicting general L2 proficiency and vocabulary size, but not morphosyntactic knowledge. On the other hand, years of exposure stood out as the most powerful predictor of the L2ers' ability to distinguish grammatical from ungrammatical conditions. With a minimum of eight years of exposure, the L2ers in this study managed to show improvements in their general L2 proficiency, vocabulary size, and ability to distinguish between grammatical and ungrammatical conditions. This length of exposure is not far from what was found in previous research. According to Muñoz's (2011) review of the literature, a minimum of ten years or even longer is required to show a positive effect in the long run. On the other hand, a length of three to five years of exposure was not shown to have a significant effect on morphosyntax (see Johnson & Newport, 1989).

Conclusion:

In conclusion, this study has shown that the length of exposure to a foreign language is far more important than when or where it has been learned and that a minimum of eight years is capable of showing improvement in the long run. However, the benefits of study-abroad should not be completely overlooked because it has been shown that it positively interacts with increased years of exposure. From a practical perspective, we have seen how this interaction is reflected positively in the higher L2 proficiency and vocabulary growth of the L2ers. Further research is needed to verify the outcome of this study especially to compare the short-term and long-terms effects of study-abroad.

References:

- Al-Thubaiti, K. A. (2010). *Age effects in a minimal input setting on the acquisition of English morpho-syntactic and semantic properties by L1 speakers of Arabic*. (Unpublished PhD dissertation), Department of Language and Linguisitics, University of Essex, UK.
- Al-Thubaiti, K. A. (2014). Age of L2 learning makes no difference in instructed settings: Input matters most. In K. M. Bailey & R. M. Damerow (eds), *Teaching and Learning English in the Arabic-Speaking World* (pp. 162-177). New York: Taylor & Francis/Routledge.
- Baayen, R. H. (2008). *Analyzing linguistics data: a practical introduction to statistics*. Cambridge: Cambridge University Press.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.
- Bialystok, E., & Miller, B. (1999). The problem of age in second-language acquisition: Influences from language, structure, and task. *Bilingualism: Language and Cognition, 2*(2), 127-145.
- Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language, 44*(2), 235-249.
- Breheny, P., & Burchett, W. (2017). Visualization of regression models using visreg. *The R Journal, 9*(2), 56-71.
- Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition, 26*(2), 227-248.

- Collentine, J. (2009). Study abroad research: Findings, implications, and future directions. In M. H. Long & C. J. Doughty (eds), *The handbook of language teaching* (pp. 218–233). Oxford: Wiley-Blackwell.
- Cunnings, I., & Finlayson, I. (2015). Mixed effects modeling and longitudinal data analysis. In L. Plonsky (ed.), *Advancing quantitative methods in second language research* (pp. 159-181). New York: Routledge.
- DeKeyser, R. M. (1991). Foreign language development during a semester abroad. In B. Freed (ed.), *Foreign language acquisition research and the classroom* (Vol. 104119). MA: D. C. Heath: Lexington.
- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*(4), 499-533.
- DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (eds), *The handbook of second language acquisition* (pp. 313-348). Malden, Mass.: Blackwell.
- DeKeyser, R. M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning, 63*, 52-67.
- García-Mayo, M. d. P., & Lecumberri, M. L. G. a. (eds). (2003). *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters
- Howard, M. (2006). The expression of number and person through verb morphology in advanced French interlanguage. *IRAL–International Review of Applied Linguistics in Language Teaching, 44*(1), 1-22.

- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99.
- Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: the status of subjacency in the acquisition of a second language. *Cognition, 39*(3), 215-258.
- Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research, 24*(1), 35-63.
- Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins Publishing Company.
- Llanes, À. (2010). *Children and adults learning English in a study abroad context.* (Unpublished PhD dissertation), University of Barcelona.
- 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, 8*(3), 189-215.
- Llanes, À., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System, 37*(3), 353-365.
- Llanes, À., & Muñoz, C. (2013). Age effects in a study abroad context: Children and adults studying abroad and at home. *Language Learning, 63*(1), 63-90.
- Muñoz, C. (2006a). The effects of age on foreign language learning: The BAF project. In C. Muñoz (ed.), *Age and the rate of foreign language learning* (pp. 1-40). Clevedon: Multilingual Matters.

- Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 Learning. *Applied Linguistics, 29*(4), 578-596.
- Muñoz, C. (2011). Input and long-term effects of starting age in foreign language learning. *IRAL-International Review of Applied Linguistics in Language Teaching, 49*(2), 113-133.
- Muñoz, C. (ed.) (2006b). *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.
- Muñoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching, 44*(01), 1-35.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle and Heinle.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nishikawa, T. (2014). Nonnativeness in near-native child L2 starters of Japanese: Age and the acquisition of relative clauses. *Applied Linguistics, 35*(4), 504-529.
- Oxford University Press, University of Cambridge, & Association of Language Testers in Europe. (2001). *Quick placement test: Paper and pen test*. Oxford: Oxford University Press.
- Patkowski, M. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning, 30*(2), 449-472.
- Pfenninger, S. E. (2014). The misunderstood variable: Age effects as a function of type of instruction. *Studies in Second Language Learning and Teaching, 4*(3), 529-556.
- Pfenninger, S. E., & Singleton, D. (2017). *Beyond age effects in instructional L2 learning: Revisiting the age factor*. Bristol: Multilingual Matters.

- R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from:
    https://www.R-project.org/
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. UK: Palgrave Macillan.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language testing, 18*(1), 55-88.
- Schütze, C., & Sprouse, J. (2013). Judgment data. In R. Podesva & D. Sharma (eds), *Research methods in linguistics* (pp. 27-50). Cambridge: Cambridge University Press.
- Segalowitz, N., Freed, B., Collentine, J., Lafford, B., Lazar, N., & Díaz-Campos, M. (2004). A comparison of Spanish second language acquisition in two different learning contexts: Study abroad and the domestic classroom. *Frontiers: The interdisciplinary journal of study abroad, 10*, 1-18.
- Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition, 26*(2), 173-199.
- White, L. (2003). *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.
- Yeni-Komshian, G. H., Flege, J. E., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean–English bilinguals. *Bilingualism: Language and Cognition, 3*(02), 131-149.

Appendix:

Here are two examples of questions which appeared in the OPT.

Part 1: In this section you must choose the word which best fits each space in the text below. For questions **6** to **10**, mark one letter **A**, **B** or **C** on your Answer Sheet.

Scotland

Scotland is the north part of the island of Great Britain. The Atlantic Ocean is on the west and the North Sea on the east. Some people **(6)** …………… Scotland speak a different language called Gaelic. There are **(7)** ………… five million people in Scotland, and Edinburgh is **(8)** ………… most famous city.

Scotland has many mountains; the highest one is called 'Ben Nevis'. In the south of Scotland, there are a lot of sheep. A long time ago, there **(9)** …………. many forests, but now there are only a **(10)** …………….

Scotland is only a small country, but it is quite beautiful.

6	**A** on	**B** in	**C** at
7	**A** about	**B** between	**C** among
8	**A** his	**B** your	**C** its
9	**A** is	**B** were	**C** was
10	**A** few	**B** little	**C** lot

Part 2: In this section you must choose the word which best fits each space in the text below. For questions **41** to **50**, mark one letter **A**, **B**, **C** or **D** on your Answer Sheet.

The tallest buildings – SKYSCRAPERS

Nowadays, skyscrapers can be found in most major cities of the world. A building which was many **(41)**…………. high was first called a skyscraper in the United States at the end of the 19ᵗʰ century, and New York has perhaps the **(42)**………… skyscraper of them all, the Empire State Building. The **(43)**……………. beneath the streets of New York is rock, **(44)** ……………. enough to take the heaviest load without sinking and is therefore well-suited to bearing the **(45)**…………. of tall buildings.

41	**A** stages	**B** steps	**C** storeys	**D** levels
42	**A** first-rate	**B** top-class	**C** well-built	**D** best-known
43	**A** dirt	**B** field	**C** ground	**D** soil
44	**A** hard	**B** stiff	**C** forceful	**D** powerful
45	**A** weight	**B** height	**C** size	**D** scale

Here are examples of the grammatical and ungrammatical test items which appeared in the GJT

Grammatical

1. The women loudly shouted across the street.

2. Susan has retired from teaching, and Wendy has too.

3. The judge is not easy to persuade on such matters.

Ungrammatical

1. *The postman had delivered always the parcels.

2. *John is happy, and Mary will soon.

3. *The manager is useful to know him sometimes.