

Efficacy of Early Study Abroad on Second-language English Morphosyntactic Competence: Effects of Age and Learning Contexts

Sangki Kim (Konkuk University)

Kim, Sangki. 2020. Efficacy of early study abroad on second-language English morphosyntactic competence: Effects of age and learning contexts. *Korean Journal of English Language and Linguistics* 20, 588–612. Early study abroad (SA) continues to be a popular form of foreign language education based on the assumed benefits of immersing oneself in the target language-speaking context. Taking a comparative perspective, this study aims to confirm an association between early SA and L2 English grammatical competence. Data were collected from questionnaires and a 72-item grammaticality judgment test administered to 195 Korean high school students, which were analyzed using ANCOVA and Multiple Regression Analysis. The results confirmed that the early study-abroad context is more advantageous than the study-at-home context for English morphosyntax acquisition. The L2 English-immersion context in the L1-speaking macro-setting was beneficial only to the study-abroad group. The findings also show that length of residence is a more influential factor than the age of arrival during early SA for improving L2 English grammatical competence. It is argued that the study-abroad context facilitated L2 English grammar more effectively than did the foreign language learning or L2 English-immersion contexts; however, the length of SA was more important than the effect of age.

Keywords: early study abroad, learning contexts, age effect, length of residence

1. Introduction

Early study abroad (SA), i.e., having school-aged or even younger children study at educational institutes in a foreign country, continues to be a popular form of educational practice. According to the Korean Educational Development Institute (KEDI 2018), 16,332 school-aged Korean students stayed in various countries (mainly English-speaking countries such as the US and Canada) for the purpose of early SA in 2018. One of the

This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

underlying reasons for early SA is the acquisition of a foreign language (Kim and Sung 2017). This benefit assumes at least two factors as key roles, namely language-learning contexts and the effect of age of acquisition. First, the SA context is widely known as distinctively different from a foreign-language-learning context and an L2 immersion context due to the qualitative and quantitative differences in authentic target language input and opportunities to use target languages. Also, sufficient empirical evidence supports the effects of age of onset on the ultimate L2 attainment (Collentine 2009, Freed 1995, Llanes and Muñoz 2013, Perez-Vidal and Juan-Garau 2011, Swain 1985).

The two factors appear to indicate obvious effects of SA on L2 learning; however, there is some complex research evidence to confidently conclude its effects. For example, age of onset appears to affect different areas of language competence. That is, while pronunciation undergoes strong age effects, morphosyntax does not appear to be susceptible of age effects (e.g., Abrahamsson and Hyltenstam 2009, Bialystok 1997, Flege, Yeni-Komshian and Liu 1999, Long 2005, Thomas and Johnson 2008). Similarly, the effects of the study-abroad learning context on improving L2 morphosyntactic competence is not conclusive (Collentine 2009, Churchill and DuFon 2006, Isabelli and Nishida 2005).

Besides these two factors, L2 attrition merits scholarly attention if we are to better understand the efficacy of early SA on L2 grammatical competence. Apparently, once early SA participants return to their L1-speaking macro-setting, they tend to experience L2 attrition (van Els 1986). This gradual loss of L2 competence is not surprising given that the returnees would scarcely use their L2 for meaningful communication in the L1-speaking macro-setting (i.e., Type 3 attrition, van Els 1986).

In this paper, therefore, I aim to examine the efficacy of early SA for acquiring L2 English grammatical competence. I first investigate context effects by comparing how the participants differed by their educational context (i.e., SA and study-at-home) and educational language (i.e., L1 Korean and L2 English) with respect to L2 English morphosyntactic competence. I then turn to examining age effects first by investigating the association between participants' early study-abroad experience with their current English morphosyntactic competence. Finally, I explore the contribution of variables of early SA to the prediction L2 English grammatical competence. The following section reviews the relevant literature.

2. Theoretical Backgrounds

2.1 Learning Contexts

Research on the effects of SA on L2 grammatical development yielded complex results (e.g., Churchill and DuFon 2006, Collentine 2009). On the one hand, the effect of SA context on L2 grammatical development has been challenged because the effect is not necessarily greater than in the study-at-home (AH) context (Collentine 2004, 2009, Cheng and Mojica-Diaz 2006, DeKeyser 1991, Lafford and Collentine 2006). Other studies, on the other hand, argued for its positive impact on L2 grammar (Kinginger 2008, Isabelli and Nishida 2005). For example, Collentine (2004) questioned the context effects. He compared an SA group of 26 participants who studied abroad for a semester in Spain to an AH group of 20 participants who took a semester-long Spanish class at university in terms of the grammatical accuracy in their oral proficiency interviews recorded before and after one semester of SA or the formal Spanish class. While showing no overall main effect of either context on grammatical accuracy, analyses of covariance (ANCOVA) revealed that the AH group outperformed the SA group in the accuracy of tense use out of five subcategories of 17 morphosyntactic measures. Therefore, he concluded that the SA experience does not improve students' overall grammatical abilities. In contrast, Kinginer (2008) showed that having one semester of SA facilitated the 23 university students' performance on the reading section in the pre- and post-test, which included items to measure grammatical competence. However, she carefully interpreted the result to infer that the SA effect on L2 French grammatical improvement is ambiguous, because approximately one-third of participants' post-test scores decreased in comparison to their pre-test results.

These conflicting findings necessitate more comparative evidence to understand the effect of the SA learning context on L2 grammatical development. Moreover, there is a dearth of research that investigates the AH context in detail. The AH context should include not only a foreign-language-learning context, but also an L2 immersion context in the L1 macro-setting. Also, most of the studies relied on college students as participants; the results, therefore, might not be applicable to young learners whose learning rate is slower than that of adults (DeKeyser 2013). Finally, most of the research studies used a one or two semester-long SA period as a treatment, which might be too short. In other

words, the treatment could be of insufficient duration to lead to a noticeable change (Churchill and DuFon 2006).

2.2 Age Factor

The age effect on L2 grammatical competence appears to be complex (Bialystok and Hakuta 1999, Bialystok and Miller 1999, Birdsong and Molis 2001, DeKeyser 2000, Flege, Yeni-Komshian and Lu 1999, Johnson and Newport 1989). Johnson and Newport (1989) supported the Critical Period Hypothesis (CPH) by demonstrating a continuing decline in L2 English morphosyntactic performance throughout the early ages of arrival (AoA) between 3 and 15. The performance did not decrease when the AoA ranged between 17 and 39. They concluded that early learners who begin learning an L2 just before the beginning of puberty tend to have native-like morphosyntactic competence. By contrast, late learners' grammatical competence is not likely to be in the native-speaker range; this holds true regardless of the length of stay in the L2 context following puberty.

Other evidence, however, suggested a different view of the effect of age on L2 morphosyntax acquisition. For instance, based on 240 Korean immigrants in the US, Flege, Yeni-Komshian and Liu (1999) reported that the overall performance of the early (*Mean* AoA = 9.7) and late groups (*Mean* AoA = 16.2) on the Grammaticality Judgment Test (GJT) did not differ significantly when confounding variables—that is, the use of L2 English and L2 Korean language and the amount of education in the US—are all controlled. In their strict replication of Johnson and Newport (1989), Birdsong and Molis (2001) analyzed the GJT performance of 61 Spanish learners of L2 English. Instead of reporting a decline up to age 16 (AoA \leq 16, $r = -0.24$, $p = .22$), they surprisingly presented a post-maturational decline in GJT performance beginning at the AoA of 17 (AoA \geq 17, $r = -0.69$, $p < .0001$).

These counter CPH studies were challenged by Long (2005), who cited, for example, a study by Flege, Yeni-Komshian and Liu (1999) on invalid measures, using individual subjects' performance falling within two standard deviations of a native mean, and Birdsong and Molis' (2001) study on the ceiling effect of non-native subjects resulting from ineffective test instruments as well as inappropriate L1-L2 pairing. Long (1996, 2005) noted that the phase of the loss of the capacity for L2 learning occurs gradually rather than abruptly. He also suggested multiple offsets for a maturational constraint from one linguistic domain to another, cumulatively at the age of six for the acquisition of a native—

like accent and in the mid-teens for the remaining domains of language, including morphosyntax.

These complex views on age effects fail to provide definitive evidence for the efficacy of early SA. Further, these investigations tend to focus on adult immigrant populations whose period of residence in an L2 context was at least more than a decade—for example, the mean length of residence (LoR) was around 34 years for DeKeyser (2000); 20 and 30 years for Hakuta, Bialystok and Wiley (2003); 10 years for Johnson and Newport (1989); 13 to 20 years for Flege, Yeni-Komshian and Liu (1999); and around 10 to 12 years for Birdsong and Molis (2001). These studies disregarded LoR in the investigation of the age effects (Birdsong and Molis 2001, Johnson and Newport 1989, p. 82, Stevens 2006) because LoR did not seem to play a significant role when it was more than 5 to 10 years (Long 2005, p. 308, Stevens 2006, p. 682). DeKeyser (2000) suggested that L2 learners with a relatively small LoR will not reach their ultimate level of attainment (p. 504). This is precisely the reason that immigrant population-based studies cannot resolve whether or not the age factor is important for early SA participants whose LoR is relatively shorter than that of immigrant participants.

2.3 Second Language Attrition

L2 attrition refers to an individual's gradual loss of proficiency—a process is recognized as being normal. A few studies have investigated the relationship between returnees' linguistic competence and increasing the period following their return from SA. Tomiyama (2000) reported that rapid loss occurs at the beginning of their return followed by a plateau; and the reverse pattern is also observed (p. 307).

Grammar was determined to resist attrition more than L2 lexicons (Weltens, Van Els and Schils, 1989). Tomiyama (2000) provided thick descriptions of L2 attrition occurring over 33 months following the return of a Japanese male child returnee at the age of eight after staying in the US for seven years. Substantial morphological change did not occur during the 19 months following his return. Although some attrition in syntax appeared after 20 months, substantial L2 attrition in morphology and syntax still was not observed for up to 33 months. The author thus suggested that L2 speakers' morphology and syntax at high L2 proficiency are not likely to be susceptible to L2 attrition. In contrast, other studies confirmed that L2 speakers with lower proficiency are more vulnerable to attrition in L2 grammar (e.g., Hansen, 2001). In sum, the literature clearly suggests that L2 attrition is at

work for returnees. Thus, this factor needs to be considered in order to precisely investigate the efficacy of early SA.

2.4 Research Purpose and Questions

The purpose of this research is to examine the efficacy of early SA for young English learners' grammatical competence. The study first explores the role of learning contexts, and then it investigates whether or not the onset age of early SA is a significant factor in improving L2 English grammatical competence. Specifically, the study investigates the following two research questions:

1. Are the three learning contexts (SA, foreign language learning, and L2 English immersion) associated with performance on the English GJT?
2. Do the AoA, LoR, and period since return (PsR) have an observable effect on performance on the English GJT among the returnees? If so, which variable is the more important predictor of the English GJT performance?

3. Methodology

3.1 Study Design and Participants

A total of 195 participants (female = 108, average age = 13.34; $SD = 1.51$) were recruited from two L2 English immersion high schools out of about a dozen, and from two public high schools in the Seoul metropolitan area in South Korea. Exclusions were applied to the analysis for the first research question: 7 participants who did not complete the English GJT; 4 who did not provide information regarding the amount of reading on the questionnaire; and 4 who were identified as multivariate outliers. These exclusions resulted in a total of 180 (female = 103, average age = 13.62, $SD = 1.59$) participants who completed the GJT and questionnaire for Research Question 1.

The analysis for Research Question 2 focused on the returnees: $n = 65$, out of the 195 participants. Exclusions were the following: 4 participants for missing data about the arrival or the departure date to or from the English-speaking country; and 2 participants for incomplete GJT results. Additionally, 7 participants did not meet the criterion for being an

SA participant. To account for these 7, Korean students' length of SA was typically less than five years (*Mean LoR* = 3.25 years for KEDI, 2005, p. 53; *LoR* = less than five years [80.8 percent of 248 participants] according to Ministry of Education, Science and Technology [MEST], 2007, p. 311). In this study, 5 participants were born in the US, and each had more than 13 years of residence in the US, these participants' language history suggests that their status can be categorized as infant bilinguals rather than SA participants for L2 learning (Baker, 2011, p. 3). Additionally, 1 participant had made multiple entries to an English-speaking country with incomplete departure dates. The remaining participant of the 7 stayed in an English-speaking country for less than one month, which clearly does not qualify as early SA. The final group of study participants was thus reduced by 13 to 52.

3.2 Data Collection and Materials

This study employed a cross-sectional design using a 72-item GJT to measure participants' L2 grammatical competence in 11 areas of morphosyntactic knowledge, together with a questionnaire to collect data on demographic characteristics, SA experience, and current English use. High school students were recruited from June to October in 2010, with assistance from four teachers working at four different high schools. Participants were instructed to make a grammaticality judgment after listening to each of the 72 items. A classroom computer presented the pre-recorded sentences to students, and the test took approximately 15 to 20 minutes to complete.

The test items were made following the model of Johnson and Newport (1989) to investigate eight types of English syntactic (determiners, pronominalization, particle movement, subcategorization, auxiliaries, yes/no questions, wh-questions, and word order) and four types of morphological (past tense, plural, third-person singular, and present progressive) competence (Appendix 1). Audio stimuli were recorded by a male English L1 speaker with a doctoral degree in English education. The stimuli were presented in a semi-random order, avoiding the consecutive appearance of sentences per pair. During the recording, each sentence was read twice with a two-second pause between repetitions. The number of the test items was reduced to 72 based on Bialystok and Hakuta's (1999) criticism of Johnson and Newport (1989) for the length of their original 276-item GJT, which requires participants' "mental vigor" (p. 70). The GJT for this study was specifically designed to take the same amount of time as the listening and indirect-speaking part of the Korean College Scholastic Ability Test to avoid an overwhelming fatigue effect for the

adolescent participants. Accordingly, the test items reliably measured participants' grammatical competence (Cronbach's Alpha = .88).

After the GJT, a pencil-and-paper questionnaire was administered to participants to collect information on the following aspects of their early SA experiences: their dates of arrival and departure to and from English-speaking countries; their recent educational language since the third year of middle school (age 12 to 13); personality; motivation; attitude toward the host country; daily English use; date of birth; and socio-economic background. The questionnaire was mostly made up of multiple-choice items. The participants were instructed to give information to the best of their knowledge; they completed the questionnaire in less than 20 minutes.

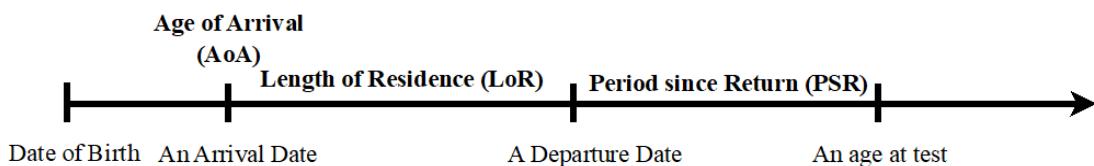
3.3 Analytical Method

For Research Question 1, a 2-by-2 between-subjects analysis of covariance (ANCOVA) was performed in order to obtain a purer effect of the two independent variables. Tabachinick and Fidell (2012) note that ANCOVA increases the sensitivity of the test of main effects and interactions by reducing the error term by adjusting dependent variable (DV) scores for differences associated with one or more covariates (CVs) (p. 197). Performance on the GJT served as the DV with increasingly high scores indicating higher L2 English grammatical competence. The two independent variables (IVs), factorially combined, are: (1) for SA with two levels, (a) having an experience of SA and (b) having no experience of SA (in other words, study at-home); and (2) for educational language since the third year of middle school (12 to 14 years old) with two levels, (a) L2 English and (b) L1 Korean. The combination of the two IVs was operationalized as a proxy for three different learning contexts: the SA context; the foreign-language-learning context; and the L2 immersion contexts in the L1 macro-setting. Two significant covariates were considered by examining a Pearson product-moment correlation matrix and conducting a principal component analysis.

For Research Question 2, a standard multiple linear regression was employed between the English GJT scores (DV) and AoA, and LoR as IVs. PsR was not entered into the model because the correlation with GJT was not significant ($p = .166$). All the statistical analyses were conducted using SPSS Statistics 20.0.

3.4 Variables

This study included several independent variables. Figure 1 below is a graphic depiction of the major time-related variables of a hypothetical participant in the study. First, AoA is understood as the age at which a participant was immersed in the target-language context (Birdsong and Molis 2001, Johnson and Newport 1989). Stevens (2006) commented that AoA is a more precise measure of the age of onset of learning an L2 than the age at which lessons began due to qualitatively different participation in the L2 context (p. 681). The AoA was calculated by subtracting a participant's date of birth from his/her arrival date. Second, LoR is operationalized as a measure of the amount of time participants were immersed in an English-speaking context (Birdsong and Molis 2001, Johnson and Newport 1989, Stevens 2006). This was calculated by subtracting the participant's arrival date from his/her departure date. Third, PsR refers to the amount of time a returnee spent in the L1 context after his or her return from SA in an English-speaking country. This was computed by subtracting the participant's age at the time of departure from his/her age at the time of testing.



Note. The bold-faced type indicates time-related major variables.

Figure 1. Major Variables

Educational language refers to the dominant language by which participants were educated, beginning with the third year of middle school (age 12 to 14). It was used as an indicator and a more precise measure of the current use of English in the L1 context based on the following information. In L2 English immersion schools, all curricula are taught in L2 English except for Korean language and Korean history. Korean high school students receive 1,190 class hours of instruction per year according to the 2009 revision of the national curriculum of Korea (Ministry of Education [MoE], 1997). This means that the immersion schools provide students with 952 class hours of instruction in English and 238 class hours in Korean for Korean language and history per year. On the other hand, the

public Korean high school provides 136 class hours of English instruction conducted in Korean. Thus, by avoiding heavy reliance on self-reported data, this variable is considered a more precise measure of current L2 English use. The study used two additional variables: (1) English reading (self-reported amount of reading in English outside of school); and (2) grade in school (1st, 2nd, or 3rd grade). The dependent variable is the score on the GJT.

4. Results

4.1 Research Question 1

The following is a description of how relevant assumptions were examined. The GJT scores were examined for univariate outliers; all scores were found to fall within 2.5 standard deviations. However, multivariate outliers were detected, leading to the exclusion of four participants (Tabachnick and Fidell 2012, pp. 99–100). The amount of reading in English outside of school (Variable = English reading) was moderately skewed (*skewness* = $-.71$, *standard error of skewness* = $.181$). Thus, a square root transformation was made of English reading. The GJT also was found to be negatively skewed. However, the degree of skewness was moderate, so it was decided not to be transformed after examining the histogram of the GJT scores. The normality of sampling distributions was a concern because of the unequal number of samples across the groups. The largest group was seven times bigger than the smallest group, which called for random exclusion. Although the sample size in each cell was not completely equalized, normality improved significantly. Tabachnick and Fidell (2012) confirmed that a formal test of homogeneity of variance is not needed when the ratio of sample sizes is less than 4:1 (p. 232). The current study meets this criterion ($29/19 = 1.53:1$). Homogeneity of regression was examined by testing interactions between the CVs and the IVs using a customized model. As Tabachnick and Fidell (2012) suggest, the alpha criterion was adjusted for because the multiple tests performed by this method of evaluating homogeneity of regression (p. 239). The test confirmed that the interactions were not significant, so the assumption was tenable. Table 1 below shows descriptive statistics for the participants in view of the first research question.

Table 1. Descriptive Statistics for Four Groups by Study Abroad and Educational Language (n = 180/101)

Study abroad	Educational Language	N (group no.)	Age at test	GJT Mean (SD)	Range	English Reading (hr/day) (SD)
Yes	English	28 (1)	16.1 (.59)	61.25 (5.62)	26	2.32 (1.27)
	Korean	29 (2)	15.4 (.68)	56.10 (6.52)	24	1.07 (1.27)
	English	15 (3)	14.3 (.39)	53.20 (7.06)	25	1.75 (1.14)
	Korean	29 (4)	13.5 (.56)	50.48 (9.80)	31	1.06 (1.09)
No	Korean	⁰ 108 (4)	16.6 (.61)	50.34(10.10)	34	1.14 (1.02)

Note. Superscripted O refers to the original sample size before random deletion; SD = Standard Deviation; GJT = the Grammatical Judgment Test scores (Max = 72).

The following procedures were conducted to determine which covariates to include in the ANCOVA. First, the Pearson correlations were examined between a number of variables and the GJT scores (DV) in Table 2 in order to confirm the variables associated with the GJT scores other than the two IVs, i.e., covariates. Of the nine variables collected from the survey, the correlation of six variables—grade, educational language, gender, SA, 3rd language learning and 3rd language use, and amount of reading in English—had a GJT of above .3, which accounts for 10 percent of the variance of GJT scores. Of these six variables, the amount of reading in English after school and participants' grade level (i.e., 1st, 2nd, 3rd,) in high school were found to be the most important covariates for the GJT scores based on the Pearson correlation matrix. The correlation between two CVs was .244; thus, no multicollinearity or singularity was detected.

There is no theoretical basis for expecting the variable of gender to be significantly correlated with the L2 grammatical performance; however, it was indeed found to be positively correlated ($r = -.392$, $p < .05$). Further examination revealed that the correlation between gender and GJT score is found to be confounded by other variables such as SA, grade, and educational language, as shown in Table 3, because those variables are significantly correlated with the GJT scores.

Before conducting an ANCOVA, a principal component analysis (PCA) with varimax rotation was performed on 10 variables, both from the GJT scores and the survey, in order to confirm the underlying processes that produced the correlations among variables. The PCA excluded 16 survey participants who did not answer the questions about their experience of learning a third language (three cases for the Thridlang_lrн variable), their use of a third language (four cases for Thridlang_use variable), and the amount of time spent using English at a private academy (nine cases for Engprivt variable).

Table 2. Correlations among the Primary Variables (n = 180)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
GJT scores (1)	1									
School (2)	.017	1								
Grade (3)	.535*	.310*	1							
	—	.511*	—	1						
Edulang (4)	.306*		.117							
	.392*	.094	.366*	—	1					
Sex (5)				.116						
Study abroad (6)	—	.079	—	.403*	—	1				
	.385*		.218*		.009					
	—	.105	—	.103	—	—	1			
3rdlang_lrn (7)	.196*		.205*		.269*	.185*				
	—	—	—	.106	—	—	.585*	1		
3rdlang_use (8)	.307*	.038	.313*		.319*	.142				
	.426*	—	.244*	—	.241*	—	—	—	1	
Engreading (9)		.129		.310*		.309*	.122	.231*		
	.184*	.220*	.174*	—	.036	—	.066	.025	.213*	1
Engprivt (10)				.077		.230*				

Note. GJT scores = the 72-item Grammatical Judgment Test result (Max=72); school = participants' school; edulang = the dominant language of recent education; study abroad = participants' experience of study abroad; 3rdlang_lrn = experience of studying another second language other than English; 3rdlang_use = whether or not participants speak another second language other than English; Engreading = the amount of reading in English after school; Engprivt = the amount of using English at a private academy. * $p < 0.05$ (2-tailed).

Table 3. Cross-tabulation for Gender

Study abroad	Male	Female	Total
Study Abroad	24	33	57
Study At-Home	53	70	123
Total	77	103	180
Grade (in school)			
First year	30	7	37
Second year	45	91	136
Third year	2	5	7
Total	77	103	180
Educational language			
English	14	29	43
Korean	63	74	137
Total	77	103	180

Four factors showed Eigen values of over 1.00. The scree plot also confirmed that a four-factor solution was appropriate. These factors accounted for 70.7 percent of the variance. The loadings for each of the variables on the four factors are shown below in Table 4. The asterisks in the table indicate loadings above .30, and the bold font indicates the highest loading for each variable. Communalities (h^2) indicate the total proportion of variance accounted for by the four factors in each variable. The GJT scores' h^2 is .839, so the four factors can be confirmed to account for 83.9 percent of the variance in that variable. The bottom row of the table shows the overall proportion of variance accounted for by each factor. Note that most of variables, which reflect participants' L2 English use, load most heavily on Factor 1. Both variables, school and educational language, load on Factor 3. This is reasonable since two of the four high schools were L2 English immersion schools. When the variable, school, was excluded, educational language loaded on Factor 1 (factor loading = $-.655$, $h^2 = .430$). Thus, the PCA provided an operational definition for L2 English grammatical competence by including 6 variables out of 11. In sum, the Pearson correlations and the PCA results suggested two variables, English reading and grade, as the most important covariates for the variance of the GJT scores (DV).

Table 4. Principal Component Loadings after Varimax Rotation (n = 164)

Variables	Factor				h^2
	1	2	3	4	
GJT scores	0.777*	-0.284	-0.060	0.022	0.839
English reading	0.672*	-0.123	-0.251	0.152	0.650
Study abroad	-0.670*	-0.351*	0.284	0.188	0.773
Grade in school	0.659*	-0.348*	0.296	-0.082	0.931
English at private academy	0.494*	0.197	0.246	-0.253	0.688
Third language use	-0.182	0.864*	-0.033	0.059	0.755
Third language learning	-0.008	0.860*	0.114	-0.041	0.784
School	0.136	0.000	0.901*	-0.096	0.689
Educational language	-0.383*	0.084	0.770*	0.165	0.553
Character	-0.021	0.035	0.027	0.964*	0.407
proportion of variance	0.238	0.187	0.171	0.110	0.707

Note. [bold] highest loading for each variable * loadings above .300.

Table 5 below shows the results of the ANCOVA, observing TYPE II Sum of Squares as appropriate for non-experimental research design with unequal cell sizes (Tabachinick and Fidell 2012, p. 219). After adjustment by covariates (i.e., grade in high school and the amount of English reading), the GJT scores varied significantly depending on whether the participants had the experience of SA, with $F(1, 95) = 11.36$, $p = .001$. The relationship

between the adjusted GJT scores and SA was of medium strength, with partial Eta Square = .107 according to the well-established criterion (Tabachinick and Fidell 2012, p. 55). The main effect of educational languages was found to be marginally significant, with $F(1, 95) = 4.630, p = .058$. For the variable of educational language, the effect size found small with partial Eta Square = .037. The interaction was not statistically significant, with partial Eta Square = .026. However, pairwise comparisons yielded some interesting results. Figure 2 graphically reveals the effects of the two IVs and the possible interaction between those two IVs on the GJT scores.

Table 5. Analysis of Covariance of the L2 English Grammatical Performance (n = 101)

Source	SS	df	MS	F	p	Partial	
						eta sq	Power
Study abroad	557.033	1	557.033	11.363	0.001	0.107	0.916
Educational language	180.693	1	180.693	3.686	0.058	0.037	0.476
Interaction	126.548	1	126.548	2.582	0.111	0.026	0.356
Covariates							
English reading	331.962	1	33.1962	6.772	.011	0.067	0.731
Grade in school	303.282	1	303.282	6.187	.015	0.061	0.692
Error	4656.995	95	49.021				

Note. R Squared = .351 (Adjusted R Squared = .317).

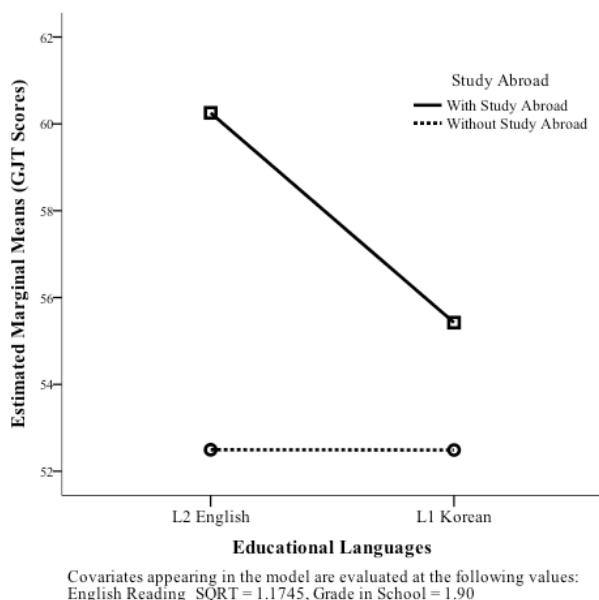
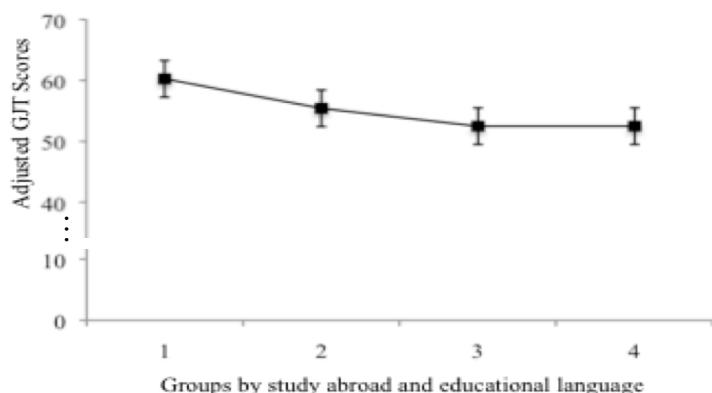
**Figure 2. Estimated Marginal Means of the Grammaticality Judgment Test Scores**

Figure 2 below suggests that educational languages have stronger associations with the GJT for the SA group, although they do not have a similar effect for the AH group. To further investigate the possible interaction effect, the study compared the adjusted means with 95% confidence interval with the observed means presented in Table 6 and in Figure 3. Further, the comparison between the observed means and adjusted means provides evidence of the methodological benefit of considering covariates.

Table 6. Adjusted Mean in the GJT Scores by Study Abroad and Educational Language

Study abroad	Educational language	Group	Observed mean	Adjusted mean	Std. error	95% Confidence interval
Study abroad	English	1	61.25	60.26a	1.382	57.511 62.998
	Korean	2	56.10	55.42a	1.329	52.785 58.062
Study at Home	English	3	53.20	52.49a	1.817	48.887 56.101
	Korean	4	50.48	52.49a	1.396	49.717 55.260

a Covariates appearing in the model are evaluated at the following values: grade = 1.90, English reading_Square Root = 1.1745.



Note. Group 1 = Study Abroad with L2 English; Group 2 = Study Abroad with Korean; Group 3 = Study At Home with English; Group 4 = Study At Home with Korean.

Figure 3. Adjusted Means with 95% Confidence Intervals for Four Groups

To confirm the differences between the groups, pairwise comparisons of each main effect were conducted using the Bonferroni adjustment and are shown in Table 7 below. For the SA groups, the educational languages were associated with a statistically significant difference in the GJT scores (see Group 1 – Group 2). Also, for those who use English as their educational language, having the experience of SA was associated with a

significant difference in the GJT scores (see Group 1 – Group 3). In contrast, the educational languages were not associated with a statistically significant difference in the GJT for the AH group (see Group 3 – Group 4). Also, for those who use Korean as their educational language, the experience of SA was not statistically associated with the GJT (see Group 2 – Group 4). The pairwise comparisons therefore suggest some interaction effect between having the experience of SA and studying in an L2 English-immersion context in the L1 macro-setting.

Table 7. Comparisons of Adjusted Mean Differences in GJT

Comparison (Group Number – Group Number)		Adjusted mean difference	Std. Error	Sig	95% Confidence interval for difference	
Study abroad	English – Korean (1–2)	4.831*	1.930	0.014	1.000	8.662
					–	
Study at home	English – Korean (3–4)	0.006	2.329	0.998	4.618	4.630
English	Study Abroad – Study At Home (1–3)	7.760*	2.268	0.001	3.257	12.263
	Study Abroad – Study At Home (2–4)	2.935	1.965	0.139	0.967	6.837

Note. Group 1 = Study Abroad with L2 English; Group 2 = Study Abroad with Korean; Group 3 = Study At Home with English; Group 4 = Study At Home with Korean

* *p*-values are adjusted using the Bonferroni method.

4.2 Research Question 2

The results of the evaluation of assumptions led to transformation of the GJT score and LoR variables to reduce the number of outliers and skewness, thus improving the normality, linearity, and homoscedasticity of residuals. A square root transformation was applied to LoR because it was negatively skewed. Two cases from the GJT score variable approached 3 standard deviations (*z* score = -2.91018). As a result, their raw scores increased by six points and reduced the amount of deviance (*z* score = -1.92979) following Tabachinick and Fidell (2012, p. 77). Table 8 presents descriptive statistics on the concluding data set for Research Question 2, and Table 9 shows the correlations between the variables.

Table 8. Descriptive Statistics for the Returnees (n = 52)

	LoR (yr)	AoA (yr)	PsR (yr)	GJT scores	*GJT scores	Original Mean age at testing (yr)
Mean	3.61	9.17	3.76	58.62	58.85	15.76
Std. deviation	3.14	4.01	2.49	5.90	5.28	0.84
Skewness	1.78	-1.05	0.75	-1.23	-0.78	-0.10
Std. error of skewness	.33	.33	.33	.33	.33	.33
Kurtosis	2.85	0.29	-0.49	1.46	-0.24	-0.41
Std. error of kurtosis	.65	.65	.65	.65	.65	.65
Range	13.25	15.26	9.26	26.00	20	3.96

Note. LoR = Length of Residence; LoR_sqrt = transformed with square root LoR; AoA = Age of Arrival; GJT = Grammaticality Judgment Test scores (Max = 72); *GJT = two outliers' raw score changed; LoR was transformed using a square root.

Table 9. Correlations between the Variables (n = 52)

	(1)	(2)	(3)	(4)
GJT scores (1)	1			
AoA (2)	-0.34*	1		
LoR (sq.root) (3)	0.47*	-0.71*	1	
PsR (4)	-0.20	-0.38*	-0.22	1

Note. * = $p < .05$.

The results of the standard multiple regression analysis, summarized in Table 10, were statistically significant, $F(2, 49) = 6.911$, $p = .002$, with R^2 at .22. The adjusted R^2 indicates that about one-fifth of the performance variability on the L2 English GJT is predicted by LoR and AoA. The regression coefficient of LoR differed from zero with a 95% confidence interval, from .720 to 5.910, whereas the regression coefficient of AoA was not significant, with 95% confidence interval, from -.493 to .454. LoR accounted for 10 percent of the variability in the GJT performance. In combination, these two IVs accounted for another .12 in shared variability. Altogether, 22 percent of the variability in the GJT scores was predicted by these two variables.

Table 10. Standard Multiple Regression of LoR and AoA on GJT scores

Variables	b	SE b	<i>sr2</i>
Constant	53.200		
AoA	-0.019	-0.015	0.00
LoR	3.315*	0.459	0.10
		$R^2 = .22$	
		Adjusted $R^2 = .19$	
		$R = .47*$	

Note. AoA = age of arrival; LoR = Length of Residence

* $p < .05$; Unique variability = .10; Shared variability = .12.

5. Discussion and Conclusion

This study began by posing two research questions regarding the effect of learning contexts and the age on the participants' GJT performance, with the aim of investigating the efficacy of early SA on the L2 English grammatical competence. For Question 1, the results confirmed that the SA context is associated with higher L2 English morphosyntactic competence. A series of pairwise comparisons indicated some interaction effect between the experience of SA and educational languages. For Question 2, the analysis demonstrated that the LoR is a statistically significant predictor of the returnees' L2 English grammatical competence, while the AoA is not.

The first part of this discussion addresses the learning contexts. This study contributed additional evidence to support the positive association between SA and higher L2 grammatical competence (Kinginger 2008, Isabelli and Nishida 2005). This result is inconsistent with some previous reports that SA is not conducive to the development of L2 morphosyntax (Cheng and Mojica-Diaz 2006, Collentine 2004, DeKeyser 1991). These studies, however, based their results on relatively short SA periods. Table 11 shows that most of the studies were based on college students whose SA lasted from four weeks to a year. Most SA programs offered by universities are typically one or two semesters long. Such a short SA experience may be insufficient to produce a noticeable improvement of L2 grammar. As Churchill and DuFon (2006) rightly observed, it is plausible that the apparent lack of effect of SA may, in fact, be the result of a period of SA.

Table 11. Study Abroad Research and Treatment

Study	Result	Study design	SA treatment
Kinginer (2008)	SA = effective, but ambiguous	SA learners' pre- and post-test SA learner vs.	One semester L2 French learning program
Cheng & Mojica-Diaz (2006)	SA = not effective	Spanish native speaker SA learner vs.	4 weeks of traditional grammar instruction (40 hours in total)
Dekeyser (1991)	SA = AH	AH learner SA learner vs.	One semester language program
Collentine (2004)	SA < AH	AH learner	2 semesters of formal classroom Spanish instruction
Isabelie & Nishida (2005)	SA > AH	SA learner vs. AH learner	One year L2 Spanish learning program

Note. SA = Study Abroad learning context; AH: At-Home learning context.

In this study, the returnees' LoR during early SA is relatively longer (*Mean* LoR = 3.61 years) than the period of SA used in the aforementioned studies. This seems to be a typical length of early SA, since KEDI (2005, p. 53) also reported that in 2005, the mean period of SA for early SA participants in Korea was 3.25 years. It therefore seems reasonable to assume that the period of early SA tends to be longer than the SA program of college students. This seems to be the reason that the current study yielded evidence supporting the facilitative role of SA-learning context for L2 grammatical improvement, while one or two semester-long college SA experiences did not.

Some interaction between SA and educational language merits attention. When the returnees were not educated in the L2 English-immersion context in the L1 macro-setting—which provides at least 952 hours of English exposure per year—their L2 English grammatical competence was significantly less than that of the returnees who were educated in the immersion context. In contrast, when the returnees were educated in Korean in the L1 macro-setting, the SA experience did not yield a significant difference in their L2 English grammatical competence between the SA group and the AH group. These group differences confirm the important role of current L2 use, which previous studies have recognized as a critical variable in adults immigrants' morphosyntax acquisition (Birdsong and Molis 2001, p. 247, Flege, Yeni-Komshian and Liu 1999, pp. 94–94).

PsR as a proxy for L2 attrition did not appear to be statistically correlated to L2 grammatical competence. While this might be because returnees' L2 grammar and receptive language skills resist attrition (Tomiyama 2000), it could be confounded by L2 learning

that took place in the AH context after participants' return. The latter seems to be more plausible given the comparison between the SA group with English and Korean as their dominant educational languages, which yielded a statistically significant difference.

The second part of this discussion argues against the age effects of early SA. The underlying rationale for choosing early SA largely comes from the Critical Period Hypothesis (CPH) based on immigrant participants. The essence of the CPH is that younger learners of an L2 outperform older learners in their level of ultimate attainment, whereas the rate of acquisition for younger learners is slower than that of older learners (Long 2005).

Previous research, which investigated whether or not AoA plays a significant role for L2 morphosyntactic competence, disregarded the effect of LoR in the L2 context (DeKeyser 2000, Hakuta, Bialystok and Wiley 2003, Johnson and Newport 1989; Birdsong and Molis 2001). They had assumed that the LoR is not a critical variable for L2 morphosyntactic learning when it exceeds five or ten years, depending on scholars (Long 2005, p. 308, Stevens 2006, p. 682).

This study set out to investigate the most effective variable for predicting the returnees' L2 English grammatical competence. In considering both AoA and LoR, results of this study contradicted results from previous research showing that LoR is a major predictor of participants' L2 English grammatical competence. This study demonstrated that the age effect was not a significant variable for L2 morphosyntactic competence when early SA participants resided abroad for a relatively short amount of time (*Mean Length of Residence [LoR] = 3.61 years*). This empirically confirms that L2 learners with a relatively small LoR do not reach their ultimate level of attainment (DeKeyser 2000, p. 504).

Based on a survey of the literature, this is the first cross-sectional study focused on adolescent speakers of L2 English who experienced SA earlier in their education, considering their recent learning contexts in the L1 macro-setting. Previous research on SA focused on college-aged participants with one or two semester-long SA experiences. Also, studies of the age effects have been conducted based on immigrant participants whose contexts for language learning and use are qualitatively different from the early SA participants. However, findings from these earlier studies have presumably been applied to questions about early SA. In response, the current study sought to examine specific evidence of the efficacy of early SA.

Nevertheless, this study has several limitations. First, it employed a cross-sectional design, which does not permit attributing causality between the IVs and the DV. However,

a vast number of studies incorporate such a design, and it has proven robust for investigating the age factor on SLA from behavioral perspectives. Additionally, no information was available concerning the onset age of English instruction, having or not having L2 English immersion education at early ages (e.g., English kindergarten), and the type of instruction received during SA, which might also influence the GJT performance. I argue, however, that, based on the previous literature, the lack of this information does not confound the results of this study. Stevens (2006), who provided methodological insights on age-related studies, observed that AoA is a more reliable measure of age of onset of L2 acquisition than the onset age of L2 instruction (p. 681). Also, previous research indicated essentially different opportunities for L2 use and learning between the L2 immersion and the L2 speaking-context (Collentine 2009, Freed 1995, Swain 1985). Thus, the influence of having English immersion education (e.g., private English kindergarten) at an early age is arguably small. Finally, Muñoz (2006) found that the proficiency or accuracy level is a stronger determinant than the method of learning—that is, naturalistic or instructional (p. 122). The results of this study provide evidence that the influence of the type of instructions given to SA participants minimally affects the GJT performance.

This study did not consider participants' language aptitude which appears to be potentially important to the L2 grammatical competence (Abrahamsson and Hyltenstam 2009). Further research is needed to more fully and precisely understand the efficacy of the increasingly popular educational practice of early SA by taking into account language aptitude.

In conclusion, this study yields valuable implications for pedagogical decisions. First, the length of SA is a stronger influence than how early one begins one's SA period in the L2 context for L2 English grammatical competence. Secondly, the advantage of the SA context actually could be mitigated by the SA group's educated language in Korean upon returning to Korea. Finally, L2 English immersion education in the L1-speaking context is more facilitative of L2 English grammar than the foreign-language-learning context only when students have SA experience.

Reference

- Abrahamsson, N. and K. Hyltenstam. 2009. The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition* 30(4),

- 481–509.
- Bakers, C. 2011. *Foundations of Bilingual Education and Bilingualism*. Bristol, UK: Multilingual Matters
- Bialystok, E. 1997. The structure of age: In search of barriers to second language acquisition. *Second Language Research* 13(2), 116–137.
- Bialystok, E. and K. Hakuta. 1999. Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong, eds., *Second Language Acquisition and the Critical Period Hypothesis*, 161–181. Mahwah, NJ: Lawrence Erlbaum.
- Bialystok, E. and B. Miller. 1999. The problem of age in second language acquisition: Influences from language, task, and structure. *Bilingualism: Language and Cognition* 2(2), 127–145.
- Birdsong, D. and M. Molis. 2001. On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language* 44(2), 235–249.
- Cheng, A. C. and C. C. Mojica-Diaz. 2006. The effects of formal instruction and study abroad on improving proficiency: The case of the Spanish subjunctive. *Applied Language Learning* 16(1), 17–36.
- Churchill, E. and M. DuFon. 2006. Evolving threads in study abroad research. In M. DuFon and E. Churchill, eds., *Language Learners in Study Abroad Contexts*, 1–27. Clevedon: BGR: Multilingual Matters Limited.
- 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 and C. J. Catherine, eds., *The Handbook of Language Teaching* (pp. 218–233). New York: Wiley.
- DeKeyser, R. 1991. Foreign language development during a semester abroad. In B. F. Freed, eds., *Foreign Language Acquisition Research and the Classroom*, 104–119. Lexington: MA: D.C. Heath.
- DeKeyser, R. 2000. The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22(4), 493–533.
- DeKeyser, R. M. 2013. Age effects in second language learning: Stepping stones toward better understanding. *Language Learning* 63(s1), 52–67. doi:10.1111/j.1467-9922.2012.00737.x
- Flege, J. E., G. Yeni-Komshian and S. Liu. 1999. Age constraints on second-language acquisition. *Journal of Phonetics* 25, 169–186.

- Freed, B. F. (1995). Language learning and study abroad. In B. F. Freed, eds., *Second Language Acquisition in a Study Abroad Context*, 3–33. Amsterdam: Benjamins.
- Hakuta, K., Bialystok, E. and E. Wiley. 2003. Critical evidence: A Test of the critical period hypothesis for second language acquisition. *Psychological Science* 14, 31–38.
- Hansen, L. 2001. Language attrition: The fate of the start. *Annual Review of Applied Linguistics* 21, 60–73.
- Isabelli, C. and C. Nishida. 2005. *Development of the Spanish subjunctive in a nine-month study-abroad setting*. Paper presented at the Selected Proceedings of the Sixth Conference on the Acquisition of Spanish and Portuguese as First and Second Languages, Sommerville.
- Johnson, J. and E. Newport. 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.
- Kim, H. and M. Sung. 2017. *Research of citizen's perception on early study abroad* (Report No.: SB170007). n.p.: Piodi Company
- Kinginger, C. 2008. Language learning in study abroad: Case studies of Americans in France. *Modern Language Journal* 92, 1–124.
- Korean Educational Development Institute. 2005. *Research of Citizen's Perception and Actual Condition on Early Study Abroad: Focusing on Compliance of Rules*. Seoul, Korea: Ministry of Education, Science and Technology.
- Korean Educational Development Institute. 2009. *Statistical Yearbook of Education*. Seoul, Korea: Ministry of Education, Science and Technology.
- Korean Educational Development Institute. 2018. *Statistical Yearbook of Education*. Seoul, Korea: Ministry of Education, Science and Technology.
- Lafford, B. and J. Collentine. 2006. The Effects of study abroad and classroom contexts on the acquisition of Spanish as a second language: From research to application. In M. R. Salaberry and B. Lafford, eds., *The Art of teaching Spanish: Second language acquisition from research to Praxis*, 103–126. Washington, D.C.: Georgetown University Press.
- Llanes, À. and C. Muñoz. 2013. Age effects in a study abroad context: Children and adults studying abroad and at home. *Language Learning* 63(1), 63–90. doi: <https://doi.org/10.1111/j.1467-9922.2012.00731.x>
- Long, M. H. 1996. The role of the linguistic environment in second language acquisition. In W. C. Ritchie and T. K. Bhatia, eds., *Handbook of Second Language Acquisition, Vol. 2: Second Language Acquisition*, 413–468. New York: Academic.
- Long, M. H. 2005. Problems with supposed counter-evidence to the critical period

- hypothesis. *International Review of Applied Linguistics* 43, 287–317.
- Muñoz, C. 2006. Accuracy orders, rate of learning and age in morphological acquisition. In C. Muñoz, eds., *Age and the Rate of Foreign Language Learning*, 107–126. Clevedon, GBR: Multilingual Matters Limited.
- Ortega, L. 2009. *Understanding Second Language Acquisition*. London: Hodder Education.
- Pérez-Vidal, C. and M. Juan-Garau. 2011. The effect of context and input conditions on oral and written development: A study abroad perspective. *International Review of Applied Linguistics in Language Teaching* 49(2), 157–185. doi: <https://doi.org/10.1515/iral.2011.008> \t _blank
- Seidenberg, M. S. and J. D. Zevin. 2006. Connectionist models in developmental cognitive neuroscience: Critical periods and the paradox of success. In Y. Munakata and M. Johnson, eds., *Processes of Change in Brain and Cognitive Development: Attention & Performance XXI*, 585–612. Oxford: Oxford University Press.
- Stevens, G. 2006. The Age-length-onset problem in research on second language acquisition among immigrants. *Language Learning* 56(4), 671–692.
- Swain, M. 1985. Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass and C. Madden, eds., *Input in Second Language Acquisition*, 235–256. New York: Newbury House.
- Tabachnick, B. G., and L. S. Fidell. 2012. *Using multivariate statistics*, 6th ed.. Boston: Pearson.
- Ministry of Education. 1997. *The 7th Curricula: High School Curricula*. Seoul: Retrieved from <http://curri.mest.go.kr/main.jsp?idx=050101>
- Ministry of Education and Human Resources Development. 2007. *The Study for the Analysis of Types of Early Study Abroad*. Seoul: Retrieved from <http://naver.nanet.go.kr/Search DetailList.do>
- Thomas, M. S. C. and M. H. Johnson. 2008. Sensitive periods in brain development. *New Advances in Understanding Sensitive Periods in Brain Development* 17(1), 1–5.
- Tomiyama, M. 2000. Age and proficiency in L2 attrition: Data from two siblings. *Applied Linguistics* 30(2), 253–275.
- van Els, T. 1986. An overview of European research on language acquisition. In B. Weltens, K. de Bot and T. Van Els, eds., *Language Attrition in Progress*, 3–18. Dordrecht: Foris Publications.
- Weltens, B. and T. van Els. 1986. The attrition of French as a foreign language: Interim results. In B. Weltens, K. de Bot and T. Van Els, eds., *Language Attrition in Progress*, 205–221. Dordrecht: Foris Publications.

Examples in: English

Applicable Languages: English

Applicable Level: Tertiary

Sangki, Kim, Professor
Department of English Linguistics and English Literature
Konkuk University
268 Chungwon-daero, Chungju-si,
Chungcheongbuk-do 27478, Korea
Tel: 010-4581-2808
Email: zzzzeyes@gmail.com

Received: October 01, 2020

Revised: October 21, 2020

Accepted: October 25, 2020