The Effectiveness of Study-Abroad on Second Language Learning: A Meta-Analysis

Jin-Suk Yang

The Canadian Modern Language Review / La revue canadienne des langues vivantes, Volume 72, Number 1, February / février 2016, pp. 66-94 (Article)

Published by University of Toronto Press

The Effectiveness of Study-Abroad on Second Language Learning: A Meta-Analysis

Jin-Suk Yang

Abstract: This study synthesizes the methods and findings of published research investigating second language (L2) learners' linguistic development in study-abroad (SA) contexts. Some studies have demonstrated that SA participation is beneficial to learners' L2 development, whereas others have not supported such a relationship. The first phase of this meta-analysis analyzed the methodological features of 66 studies, while the second synthesized 11 quantitative studies comparing L2 linguistic gains between SA and at home (AH) learners. Results from the comparisons indicate that SA learners outperform AH learners and that short-term SA is more effective than long-term SA in terms of L2 linguistic development. Along with outlining potential concerns about conducting meta-analyses on SA research, I conclude this article with suggestions for future research directions.

Keywords: study-abroad, research synthesis, meta-analysis, effect size

Résumé : Cette étude est une synthèse des méthodes et des résultats de travaux publiés portant sur le développement linguistique d'apprenants de langue seconde (L2) dans le contexte d'études à l'étranger. Certaines études montrent que les séjours à l'étranger profitent au développement de la L2 chez les apprenants, alors que d'autres études récusent cette relation. La première phase de cette métaanalyse est une analyse de la méthodologie de 66 études; la seconde phase fait la synthèse de 11 études quantitatives comparant les acquis linguistiques en L2 d'apprenants à l'étranger et au pays. Les résultats des comparaisons indiquent que les apprenants à l'étranger obtiennent des résultats nettement supérieurs à ceux de leurs collègues restés au pays, et que les courts séjours sont plus efficaces que les longs en ce qui concerne le développement linguistique. Je conclus sur les difficultés éventuellement associées à la métaanalyse de travaux portant sur les études à l'étranger, et sur des suggestions en vue de recherches ultérieures.

Mots-clés : études à l'étranger, synthèse de recherches, métaanalyse, ampleur de l'effet

72, 1, (February / février), 66–94 doi:10.3138/cmlr.2344

This article reports a meta-analysis of published studies that examined second language (L2) learners' linguistic development in Study-Abroad (SA) contexts. As Rehner and Mougeon (2003) observed, learners should be equipped, not only with grammatical knowledge, but also with sociolinguistic knowledge according to their situational and social environments to become advanced L2 users. L2 classroom contexts can be limited in providing language learners with opportunities to experience various L2 discourse options (Regan, Howard, & Lemée, 2009), however. Accordingly, many researchers as well as teachers and program administrators have advocated and organized SA programs, assuming they would help learners to attain higher levels of L2 proficiency (Freed, 1995). Many international tertiary-level institutions around the world now encourage language learners to take part in SA programs with the expectation that participating students will become more proficient L2 users upon their return. But does research show they really do? That is the question I set out to evaluate in the present investigation.

Recognizing that the meanings and experiences of SA vary considerably according to the ethnic origin of students, the target language, their motivation, and their ultimate goals for SA, following Kinginger (2009), I narrowed the scope of SA in the present study to "a temporary sojourn of pre-defined duration, undertaken for educational purposes" (p. 11), which I further subcategorized into four domains: (a) full study abroad for a foreign degree qualification, (b) study as part of an academic partnership within a home degree or a joint degree involving institutions at home and abroad, (c) exchange programs, and (d) private language school programs.

Over the past two decades, the popularity of SA has been pervasive in academic programs and societies and that popularity has led to policy changes at the institutional level (Roberts, Byram, Barro, Jordan, & Street, 2001) as well as to the accumulation of notable academic works that discuss the SA learning processes from a wide range of perspectives (e.g., Barron, 2003; DuFon & Churchill, 2006; Jackson, 2008; Pellegrino Aveni, 2005). Commonsense, intuitive beliefs suggest that SA participation should be helpful for learners' linguistic development because it allows for intensive, regular, contextualized L2-use opportunities in situ (Rehner & Mougeon, 2003). A closer look at individual SA studies, however, indicates contradictory findings about the relationship between SA participation and L2 linguistic development (Wang, 2010).

Therefore, this is an appropriate time to synthesize the existing SA research in order to evaluate and discuss previous findings as well as future prospects for SA research. As has been widely recognized,

research synthesis can contribute to the extension of knowledge by combining and comparing individual research studies (Dinsmore, 2006). For the present research, I adopted the methods of quantitative meta-analysis in order to provide a comprehensive, representative account of the domain in question. Qualitative research allows us to find unique patterns that might otherwise remain unexamined if one were to look at the numerical results from meta-analysis alone, including how each individual learner differently functions in relation to host communities and L2-learning opportunities. Yet, as Norris and Ortega (2007) argued, the systematic, detailed approach to calculating effect sizes in meta-analysis allows researchers to move beyond "summarizing individual studies, presenting claims put forth by selected primary researchers, or tallying statistically significant versus nonsignificant results" (p. 808).

In what follows, I first briefly review SA research strands in order to describe the scope, history, and content of the relevant SA research. Then, I report the procedure for the present meta-analysis. Along with descriptions of the methodological features of the selected SA research, I introduce the eligibility criteria for inclusion in the subsequent quantitative meta-analysis. The final section of the article documents the general effect sizes of the SA studies examined, and the varying effect sizes according to the duration of SA are discussed.

Issues in SA research

In this section, I sketch out three major issues in SA research: first, the effectiveness of SA in comparison to at-home (AH) contexts; second, the target L2 features in question (i.e., what to measure); and third, possible ways to evaluate SA experiences (i.e., how to measure). These questions remain controversial and are worthy of investigation because many researchers and educators have argued for SA experiences as a global way of achieving L2 competence.

The effectiveness of SA in relation to AH contexts

Motivated by Carroll (1967), whose work showed that experience with a target language is one of the strong factors predicting learners' L2 proficiency, a number of SA studies have focused on the effects of L2 exposure on L2 linguistic development. In general, it appears that previous studies demonstrated a positive relationship between SA participation and L2 linguistic development (Freed, 1995). For instance, Regan (1998) reiterated that constant interaction with native speakers of French was an important factor for the development of sociolinguistic competence. This positive relationship was also reported in Brecht,

Davidson, and Ginsberg's (1995) study, which documented English speakers' lexical and speaking-skill gains in Russian.

Researchers, however, were not unanimous in reporting such linguistic gains over periods of sojourn. Marriott's (1995) analysis of the use of Japanese honorifics by English speakers found that, while the learners became more sensitive to variation, they nonetheless did not reach the level of native-speaker norms in terms of appropriate honorific use. Also, Freed, Segalowitz, and Dewey (2004) found that AH learners enrolled in intensive language programs outpaced SA learners in terms of L2 fluency. These contradictory findings have led researchers to take into account such variables as learners' starting proficiency and participation in formal instruction. For example, Trenchs-Parera (2009) investigated the effects of formal instruction (FI) and of SA to trace 19 Catalan/Spanish undergraduate students' English fluency development. The results indicated that FI allowed learners to have extra time to think by "inserting more silent pauses and self-repetitions within the speech" (p. 386). That study lends support to DeKeyser's (2010) study, which advocated the need for learners to have foundational knowledge of L2 grammar prior to departure to enable noticeable L2 linguistic progress in SA contexts. These findings showed the value of incorporating FI and informal L2 exposure, highlighting the necessity of systematic, formal L2 instruction before SA participation.

The heart of this debate lies in expectations about SA: Do SA opportunities really provide learners with L2-immersive environments during language-learning experiences? Concerted efforts to accurately measure L2 learners' contextualized L2 opportunities began to challenge such an assumption. Ranta and Meckelborg (2013) developed a computerized language activity log (LAL), which was specifically designed to examine the quality and quantity of L2 exposure. The study showed that, despite notable individual variations in the quantity and the quality of L2 interaction, the participants in general had more exposure to receptive L2 skills (i.e., reading and listening). In a similar vein, in his longitudinal study tracing American students' Chinese fluency development, Du (2013) deployed the notion of time-on-task to examine the ways in which students' actual amount time of using Chinese influenced their fluency in that language. The results demonstrated that varying degrees of L2 community involvement affected the students, providing differential amounts of speaking time with native speakers.

L2 development in SA contexts

Vocabulary acquisition and lexis development constitute an important dimension of SA research. In particular, cognitive processing skills are a major factor in differential L2 achievement in the context of SA. O'Brien, Segalowitz, Freed, and Collentine (2007) showed that phonological memory played a role in vocabulary development and oral fluency during SA. Tokowicz, Michael, and Kroll (2004) examined working memory capacity and single-word translation errors, showing that the contextual distinction between SA and AH needs a critical appraisal, for SA students' good performance was mediated by their higher working memory capacities.

At the sentence level, the issue of how much SA students' speech production sounds native-like has received much attention. Llanes and Muñoz (2009) reported SA students' better performance in the domains of oral fluency, accuracy, and listening compared to AH students. Isabelli-García's (2010) study, on the other hand, found no difference between the two contexts, showing that spending a semester abroad and practising the target language may not necessarily facilitate the process of acquiring gender-agreement accuracy. Finally, researchers have traced how learners acquire L2 competency at the discourse level. Paige, Cohen, and Shively (2004) reported that spending time abroad in French- and Spanish-speaking countries had a positive impact on intercultural development. Similarly, Masuda (2011) supported the beneficial aspects of SA in the domain of language learners' interactional competence.

In recent years, SA research has celebrated its social turn (Block, 2003) by incorporating multidisciplinary perspectives on issues such as identity and feminism (Norton, 2000; Polanyi, 1995; Siegal, 1995), language socialization (Duff, 2003; Wang, 2010), and Vygotskian sociocultural theory (Jackson, 2008; Yang & Kim, 2011). Mindful of the social categories at work, research within the sociological tradition attempts to depict SA as a profoundly gendered experience. Polanyi (1995), for instance, problematized the situation in which male students exhibited higher L2 listening achievement than did female counterparts. By adopting qualitative research methods, she revealed that the female students' discomfort and frustration in the face of sexual harassment during their SA period influenced not only their learning opportunities but also their ultimate L2 attainment. Similarly, Siegal (1995) documented how her participants deliberately articulated erroneous L2 pragmatic forms when conflicts arose between gender-sensitive pragmatic demands and perceived identity. In contrast, Churchill's (2009) case study showed how a learner's gender as male

contributed to the creation and the maintenance of contextualized L2 opportunities through manoeuvring around conversational topics.

From self-report to in-depth interview: How to measure gains in SA experience

Reporting gains through SA in a systematic, unbiased way is important for providing convincing evidence regarding the effectiveness of SA. In the earlier model of SA studies, learners were asked to self-report their gains and the data were used to support the efficacy of SA. This practice, in spite of its contribution to the field, had not been without limitations, partly due to its lack of validity (Collentine, 2009). Researchers of linguistically oriented study-abroad research, thus, have been encouraged to adopt a standardized testing instrument and to assess the participants' L2 development (e.g., Hernández, 2010; Yashima, Zenuk-Nishide, & Shimizu, 2004).

Those with a specific L2 feature in mind, however, ought to develop a test or questionnaire, which needs to be proven statistically valid prior to research. Mora and Valls-Ferrer's (2012) study examined the relationship between fluency, accuracy, and complexity in English learning among Catalan-speakers during SA. They recorded the students' speech over a two-year period at three different phases and compared their speech samples with those of native English speakers. Cohen and Shively (2007) revealed that SA students' pragmatic ability to make a request and an apology developed toward the norm of native speakers. Assessing speech production in relation to that of native speakers is helpful in that it can reveal the extent to which learners show improvement in the acquisition of particular linguistic items. However, we are confronted with the fact that these data may not be adequate for secondary data processing such as in meta-analysis, for those results may lack statistical validation due to small sampling and the different operationalization of a construct. Also, the question of the concept of *native speaker* as the ideal model of language users (Kramsch & Whiteside, 2007) provides another complication of this type of research.

Along with questionnaires and statistics, the use of qualitative data as an additional source gained momentum in the 2000s. Cheng and Fox (2008) focused on interview data to determine what could contribute to successful academic acculturation. The analysis of 56 international students' daily L2 experiences in Canadian universities indicated that learners' preparedness in terms of language proficiency before departure as well as in relation to the host institution's academic and social support were critical for successful academic engagement. This finding is in line with Dewey, Bown, and Eggett (2012),

who analyzed students' responses to a SA social interaction questionnaire and argued that the varying degrees of involvement in L2 communities often determine success or failure in the development of social networks with native speakers, thus affecting language-use opportunities. Stewart (2010) compared the results of pre-tests with those of post-tests and analyzed students' journal entries to understand how students' language gains related to the development of social identities in the SA contexts.

As Kinginger (2009) noted, SA researchers have dealt with the following questions: "How much competence do study abroad participants bring back with them, and is it different in kind from the competence that classroom learners develop? What is the optimum length of study abroad programs?" (pp. 29–30). In light of these questions, while previous reviews of qualitative research (e.g., Kinginger, 2009; Wang, 2010) have provided a general picture of SA-scholarship development, issues remain unresolved with respect to systematicity in choosing the literature reviewed as well as validity in reporting empirical results based on null hypothesis statistical testing (NHST) (Oswald & Plonsky, 2010). For these reasons, it is imperative to obtain a comprehensive, coherent synthesis of SA research. To this end, the present study employs quantitative meta-analysis methods. Keck, Iberri-Shea, Tracy-Ventura, and Wa-Mbaleka (2006) suggested that a quantitative research synthesis allows for systematic findings through the standardization of results across studies and comparison of effect sizes. To my knowledge, no quantitative synthesis of SA research has yet been published. Thus, the present research may initiate a new discussion of the effectiveness of SA. For the present study, three research questions were formulated:

1. In terms of methodology and target L2 features, what kinds of SA studies have been reported to date?
2. How effective is SA in general for L2 linguistic development?
3. Does the overall length of residence in SA settings make a difference in promoting L2 linguistic development?

Method

Data collection and analysis

For this meta-analysis, data were collected through the following online databases: *ERIC, FRANCIS, Linguistics & Language Behavior Abstracts, ProQuest Educational Journals, Scopus,* and *JSTOR* (see In'nami & Koizumi, 2010 for popular databases in research synthesis). I did not specify a timeline for relevant SA studies, hoping to fulfil the

purpose of examining historical developments in this domain. Using keyword searches for data collection, I typed "study-abroad & second language learning" in Abstracts, within the domain of "peer-reviewed" and "scholarly journals." Also, I manually consulted *Frontiers: The Interdisciplinary Journal of Study Abroad*, which deals exclusively with SA issues, and included studies focusing on learners' L2 development. Following Norris and Ortega (2000) and Russell and Spada (2006), I excluded unpublished research papers, theses, dissertations, research reports available only for a fee, and conference presentations as well in order to make the meta-analysis readily replicable in the future and to ensure the quality of the research. Given the author's linguistic and material resources, only publications written in English were considered. This search process resulted in the identification of 66 potentially relevant published studies.

The analysis of these primary studies took place in two phases (*cf.* Russell and Spada, 2006). First, all 66 studies in the database were coded for a wide range of theoretical and methodological characteristics. The purpose of this phase was to organize the data in preparation for the quantitative meta-analysis and, in the process, to identify methodological/target characteristics which had been examined.

Findings

Research Question 1: In terms of methodology and target L2 features, what kinds of SA studies have been reported to date?

Each of the 66 studies was coded according to a wide range of features (presented in Appendix A). Specifically, I considered publication year, research design, the participants' first and target languages, target features, and length of SA residence.

Table 1 shows the years of publication of the 66 SA studies. The earliest SA study in the database search appeared in 1991. Most subsequent studies were published in the 2000s, particularly from 2006 to 2010. This trend suggests that SA research is a relatively young field of inquiry, warranting more academic attention (Collentine, 2009).

Table 1: Publications by year

Year of Publication	No. of Publications	(%)
1991–1995	1	1.5
1996–2000	4	6.2
2001–2005	12	18.5
2006–2010	30	44.6
2011–present	19	29.2

Table 2: Publications by type of method and target feature

Type of Methods			(%)
Quantitative	Linguistic	23	
	Social	4	31
	L+S	4	
Qualitative	Linguistic	7	24
	Social	17	
Mixed-method			7
Review			4

Table 2 indicates the distribution of SA research in terms of type of research methods used and target L2 features. Of 66 studies, 31 adopted quantitative research methods, using either descriptive or inferential statistics. In addition, 24 qualitative studies used various research methods ranging from interviews to stimulated recalls, journal writing, and observation.[1] Seven studies used complementary mixed methods to provide a comprehensive perspective on L2 development during sojourn periods. Finally, four studies provided a historical overview of SA research.

Thirty studies reported the development of L2 linguistic features in SA learners. The primary purpose of inquiry in this strand was to investigate SA learners' development of L2 skills (i.e., reading, listening, speaking, and writing) in varying domains of phonological awareness (O'Brien et al., 2007), vocabulary acquisition (Dewey, 2008), morphosyntactic development (Marqués-Pascual, 2011), pragmatic proficiency (Taguchi, 2011), and interactional competence (Masuda, 2011), to name a few. The studies demonstrated SA learners' linguistic development either in comparison with AH learners or in terms of their performance over time. While standardized tests such as IELTS or TOEFL were routinely used, researchers such as Chang (2010), Shardakova (2005), Duperron and Overstreet (2009) used hands-on assessment tools such as a discourse completion task, grammaticality judgement task, and picture-cued oral narration task.

The second strand of research was geared toward understanding the dynamics of learners' SA experience. The researchers thus documented in detail SA learners' changes in motivation (Allen, 2010), beliefs (Wilkinson, 1998), identities (Kinginger, 2013) and intercultural awareness (Paige et al., 2004) and explored how these changes could potentially affect language acquisition. This approach favoured qualitative research methods such as written reports and oral reflections (Liskin-Gasparro, 1998), e-journals (Stewart, 2010), and interviews (Allen, 2010).

Finally, four studies integrating two research methods showed the ways in which SA participation and language outcomes are shaped by the immediate sociocultural environment. For instance, Davis (2007) addressed the question of how South Korean SA students' seeming failure to follow Australian-English pragmatic norms was in fact deeply connected to the students' preferences for North American English, a desirable variant in Korean society. Taguchi's (2011) findings shed light on why language learners showed pragmatic deviation after sustained engagement with native speakers in SA contexts. She found that individuals differed in the "type of intensity of the contact, leading to different paths toward pragmatic mastery" (p. 615) and suggested that analyzing learners' pragmatic behaviour needs further understanding of the kind of relationship created in the SA contexts.

Although Larsen-Freeman (2000) observed a general emphasis on quantitative data in the field of L2 research, this result indicates that SA research is relatively well balanced between both quantitative and qualitative research orientations. It further shows that the features of the target language in question lead researchers to choose a particular research orientation. For instance, when the purpose of the study is to understand L2 learners' linguistic development, researchers tend to choose a pre- and post-test design using quantitative methods. In contrast, researchers focus on qualitative data when they want to examine the development of social aspects of language learning. It can be argued, then, that the SA field has been shaped by considerations of methodology and target features.

Table 3 shows the distribution of SA research based on target languages. Because some studies had multinational participants targeted at learning various languages (e.g., Amuzie & Winke, 2009; Cohen & Shively, 2007), these are counted more than once. English and Spanish were the two major target languages, followed by French, German, Japanese, and Russian.

Table 4 shows, in part, how English and Spanish have become the two major target languages in SA research. Note that the learners' first

Table 3: Publications by target language

Target Language	No. of Publications	(%)
English	27	41.5
French	9	13.8
German	3	4.6
Japanese	3	4.6
Russian	3	4.6
Spanish	21	32.3

Table 4: Study participants by nationality and first language

Nationality / First Language	No. of Studies	(%)
US, UK / English	32	49.2
China / Chinese	10	15.4
Korea / Korean	7	10.8
Spain / Spanish	7	10.8
Japan / Japanese	6	9.2
Others	18	27.7

languages are far more diverse than their target languages (see Appendix B for a full list of study participants by first language and nationality). The distribution of participants shows that English speakers learning Spanish and Asian-language speakers learning English have been the major focus of SA inquiry. This empirical finding supports Block's (2007) argument that the literature in SA research on language learning has reported mostly about the experiences of students from the United States and to a lesser extent from Europe and Japan.

The heavy reliance on American students in SA research may be due to the logistical challenges most international students face as they attempt to gain entry to English-speaking countries with a student visa. However, the observation that the number of Asian international students constitutes more than 80% of the entire foreign-student population of the world (Altbach & Bassett, 2004) requires that researchers carefully examine how SA opportunities may have different cultural currency across geographical and sociopolitical contexts: volitional choice on the part of American students (e.g., to experience a different culture) versus investment on that of Asian counterparts (e.g., to increase individual competitiveness by learning the Lingua Franca) (*cf*. Kinginger, 2009). That said, it appears that a body of research addressing the students' different positionalities in relation to SA warrants a better theorization.

Study eligibility criteria to be included in quantitative meta-analysis

For the second phase of the research, it was necessary to eliminate studies that were methodologically inappropriate for carrying out a meta-analysis. Inclusion and exclusion criteria were developed based on the research questions guiding the present study. In order to be included in the second phase of meta-analysis:

(1) The study had to employ original, empirical data suitable for inferential statistics with experimental control-group designs (i.e., SA versus AH contexts). Thus, studies employing small

sampling (e.g., DeKeyser, 1991) or other statistics such as VARBRUL[2] (e.g., Díaz-Campos, 2004) were excluded for the second analysis.

(2) The independent variable in the primary studies had to involve the experience of overseas residence in a target-language setting over a certain period time.
(3) The dependent variable in the primary studies had to involve the measurement of participants' linguistic performance.
(4) Given the influence of age on language learning, the study had to investigate participants of post-secondary-school age.
(5) When multiple publications were based on the same data (e.g., Freed et al., 2004), only one study was considered for the second data analysis.
(6) Some studies investigated the effectiveness of SA in comparison with AH learners as well as with *domestic intensive* learners (e.g., Dewey, 2008). In such cases, only the effect sizes from SA learners and AH learners were used in calculations.
(7) The study design had to involve a treatment/control group comparison.

Application of these guidelines yielded 11 studies suitable for the second stage of quantitative meta-analysis among the 66 SA studies originally identified. Two other graduate students majoring in applied linguistics verified the coding process in the second analysis. I found 65 effect sizes from the 11 studies, and averaged the effect sizes in each study for comparison across the studies.

Research Question 2: How effective is SA in general for L2 development?

After identifying methodologically suitable data for quantitative meta-analysis, I first calculated individual effect sizes using Cohen's *d*, which is widely used among meta-analysts (Glass, McGaw, & Smith (1981). This statistic refers to the difference between two means (i.e., an experimental and a control group) divided by a standard deviation. In general a *d* value between 0.2 and 0.5 is regarded as a small effect, a *d* value between 0.5 and 0.8 is regarded as a medium-sized effect, and a *d* value of more than 0.8 is regarded as a large effect. However, in their more "localized interpretations of effect sizes" (p. 12), Plonsky and Oswald (2014) provided the new L2 field–specific benchmarks of small ($d = 0.4$), medium ($d = 0.7$), and large ($d = 1$), since Cohen's scale may underestimate the range of effects in L2 research.

It is important to note the relationship between the value of effect sizes and directionality. The two groups in Cohen's *d* are experimental

Table 5: Descriptive statistics for effect sizes of post-test comparisons: All studies (11)

Weighted Mean	0.75
Standard Error	0.101
Minimum	−0.504
Maximum	7.797
CI (95%)	Upper: 0.951 Lower: 0.555

and control groups. Subtracting the mean of an experimental group from that of a control group yields a directional value, with a positive result representing improvement or change in the predicted direction. In contrast, if the result is negative, it represents deterioration or opposition to the predicted direction. With regard to the present study, a positive value (+) generally represents better performance of experimental (i.e., SA learners) groups. Nevertheless, caution should exercised because this interpretation does not always hold. In Lafford (2004), for instance, a negative effect size indicates greater L2 improvement of SA learners on the grounds that less use of L2 strategies is presumed to be evidence of L2 proficiency improvement. For this reason, I converted Lafford's negative effect to its absolute value (*cf*. Plonsky & Oswald, 2014). Table 5 presents the descriptive statistics for post-test comparisons of the 11 studies. The effect sizes ranged from 0.5 to 7.8, with a mean weighted effect size of 0.75. Following Plonsky and Oswald (2014), I suggest that the present study indicates a medium effect in the realm of L2 research. These data support the argument that SA experience could lead to greater L2 linguistic attainment compared to AH classroom learning.

The effect sizes for each of the 11 studies, after they were all weighted by sample size, are presented in Table 6 (see Appendix C for a full description of each study). In particular, the four studies in Figure 1 reported effect sizes of more than 0.8, suggesting that SA groups outperformed AH learners in terms of L2 linguistic post-tests after sojourn experiences.

The O'Brien et al. (2007) study, with an exceptionally strong effect size of 7.797, is worthy of special attention. According to the data, the SA learners outpaced the AH counterparts in terms of general oral ability and speech rate in oral fluidity, both of which were statistically significant at $p < .001$. Arguably, this could be attributed to the influential role of SA contexts in oral-fluency development itself. However, Vatz, Tare, Jackson, and Doughty (2013) critiqued this study, warning about the lack of consideration of the learners' L2 starting proficiency within and across the groups. They argued that "the difference in

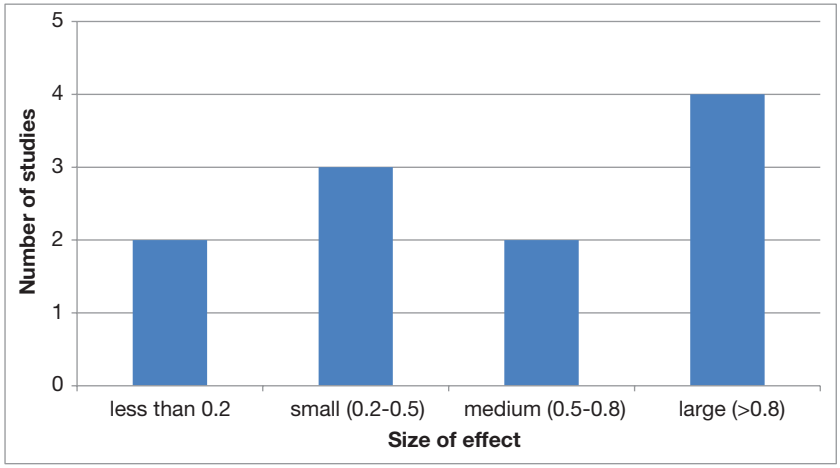

Figure 1: Effect sizes by number of studies

Table 6: Effect sizes for post-test comparisons, individual studies

	Authors	Effect Size Basis	Total *N*	Target Skills	Effect Size
1	Collentine (2004)	M & *SD*	46	Morphosyntactic and lexical proficiency	0.433
2	Dewey (2008)	M & *SD*	42	Vocabulary proficiency	5.454
3	Isabelli-García (2010)	M & *SD*	24	Grammatical knowledge	−0.504
4	Lafford (2004)	M & *SD*	46	Communication strategies	1.772
5	Marqués-Pascual (2011)	M & *SD*	40	Morphosyntactic development	0.475
6	Martinsen, et al. (2011)	ANOVA (*F* value)	76	Oral proficiency	0.538
7	O'Brien, et al. (2007)	M & *SD*	43	Phonological memory	7.797
8	Sasaki (2011)	M & *SD*	18	Writing proficiency	0.302
9	Segalowitz & Freed (2004)	M & *SD*	33	Oral proficiency	−0.054
10	Serrano,et al. (2011)	M & *SD*	62	Written and oral proficiency	0.633
11	Taguchi (2011)	M & *SD*	42	Pragmatic proficiency	0.928

starting proficiency may have interacted with the students' aptitude or learning context" (p. 238).

In fact, Segalowitz and Freed's (2004) data ($d = -0.054$) reflect the potential role of individual cognitive skills to account for differential

L2 outcomes within the SA opportunities. They argued that isolating learning context from innate individual abilities would be at the expense of inquiring into the dynamics of learner–context interactions. Thus, although they found that the SA group made greater oral gains than the AH group, they argued that these gains were not so much attributable to the amount of out-of-class contact as to individual cognitive processing abilities. This led them to conclude that the "the increased opportunities available to learners in the SA context did not necessarily result in oral performance gains over the semester" (p. 193).

A close look at the effect sizes of individual studies complicates common assumptions about SA learning. In the realm of grammatical competence, for instance, previous studies have indicated that SA students did not outperform AH learners (Collentine, 2009). A similar result is found in Isabelli-García (2010), whose effect size indicates the opposite direction (i.e., AH learners outperformed SA counterparts). In accounting for the result, she speculated that sociocultural pressure that SA learners face could have affected their sensitivity to subject–verb agreement in Spanish, for "meaning will always trump form" (p. 300) in SA contexts. That is, as the SA learners were more likely to encounter a mentor–apprentice interaction mode with native L2 speakers (Pellegrino Aveni, 2005), they may have put more emphasis on meaning as a survival strategy for successful communication. A matter of instrument validity arises here, in that her study employed a written grammaticality judgement test instead of oral data to measure grammatical accuracy. Given that both groups' L2 proficiency was assumed to be intermediate, it appears that AH learners may have benefitted from this methodological choice, for L2 instruction and classroom activities consisted of explicit rule explanations with a series of review questions. In a similar vein, Regan (1995) noted that SA learners may have difficulty in developing a uniform, standard L2 knowledge because they are more likely to be exposed to linguistic variations that sometimes do not reflect prescriptive grammar rules.

Research Question 3: Does the length of residence in SA settings make a difference in promoting L2 linguistic development?

The third research question asked whether the duration of SA residence affects L2 learners' linguistic gains. While there is no general agreement as to whether and to what extent SA residence can be regarded as short- or long-term, my review of the methodological features of the 66 studies indicates that one semester (approximately 13 weeks) can be considered as a cut-off point for the distinction. Table 7 shows the short vs long coding results for the 11 individual

Table 7: Effect sizes according to length of SA residence

SA Residence	Study	Duration	ES
Long	Isabelli-García (2010)	16 weeks	−0.504
	Martinsen, et al. (2011)	32 weeks	0.538
	Sasaki (2011)	up to 3.5 years	0.302
	Taguchi (2011)	36 weeks	0.928
			Average: 0.458 (0.375)
Short	Collentine (2004)	13 weeks	0.433
	Dewey (2008)	11 weeks	5.454
	Lafford (2004)	13 weeks	1.772
	Marqués-Pascual (2011)	13 weeks	0.475
	O'Brien, et al. (2007)	13 weeks	7.797
	Segalowitz & Freed (2004)	13 weeks	−0.054
	Serrano, et al. (2011)	13 weeks	0.981
			Average: 0.981 (2.138)

studies. Seven were coded as short-term SA (from 11 weeks up to 13 weeks), and four were coded as long-term SA (more than 14 weeks to up to 3.5 years). The average effect size for long-term SA residence is smaller than that for short-term SA residence, supporting Rees and Klapper's (2007) conclusion about the effectiveness of short-term SA.

Some cautions are necessary in interpreting the present results. First, Table 6 indicates a wide range of variation in effect sizes within each category. This serves as a reminder that effect sizes should be interpreted as a general indicator for gauging practical significance, not as a prescriptive tool (Plonsky & Oswald, 2014). I do not suggest this result be applied indiscriminately to argue against the effectiveness of long-term SA for L2 development. My second observation relates to the magnitude of effect sizes and the scope of measurement. It seems that the magnitude of effect sizes corresponds to the specificity of the L2 domain in question: The more specific the skill that a test assesses, the greater the effect size it indexes. This interaction among the scope of measurement, effect size, and length of SA duration points to the multidimensional aspects of language testing when one assesses and reports linguistic gains in SA research (Hulstijn, 2010).

Finally, although the present research was not primarily concerned with SA research from a sociological standpoint, it nevertheless offers some insights on such a perspective. Norton (2000) and Morita (2004) argued that English-language learners may not always be able to choose opportunities to speak in a new L2-learning context. This point is empirically evidenced by Ranta and Meckelborg (2013), who point out that sustained engagement in L2-rich environments may not necessarily indicate a more productive L2 use. In situations where English is regarded as a commodified resource (Heller, 2010) and the

global language (Phillipson, 1992), the ESL learners' racialized experiences and struggles for opportunities to speak in English-speaking countries may not be the same as native English speakers' SA for French or Spanish acquisition, which is mainly pursued through individuals' voluntary choice (Kinginger, 2009). Analyzing the sociopolitical dimensions of SA can therefore complement linguistically oriented SA research by situating L2 learning in a sociopolitical context that may otherwise go unnoticed. In summary, in answer to Research Question 3, I found that longer time spent overseas did not relate to greater L2 linguistic development.

Summary and implications

As Oswald and Plonsky (2010) emphasized, quantitative research synthesis can shed light on individual study results while reducing potential biases made by an author's personal, intuitive judgement. By employing a coding scheme that probed numerous features of a comprehensive, systematically gathered corpus of individual SA studies, I have presented a detailed picture of theoretical and methodological trends and issues in this domain of educational activity. First, in terms of study trends by publication year, SA research in the database first appeared in 1992, with a substantive number of studies published in the late 2000s. Second, in regard to methodological orientations, quantitative and qualitative research methods were about equally distributed. Third, analysis of the focus of SA inquiry showed that, while SA researchers are mainly interested in L2 learners' linguistic development in SA settings, attention to SA learners' changes in motivation, beliefs, and intercultural competence is gaining momentum. Fourth, the profile of study participants showed that Asian language speakers learning English and English speakers learning Spanish and French (and Japanese) were the most frequently reported participants in research studies. Fifth, the general effect size of SA experience was 0.8, which suggests medium-sized explanatory power. Finally, with regard to the relationship between length of residence and L2 gains, short-term SA residence (0.981) had more practical significance than did long-term SA residence (0.458).

Jeon and Kaya (2006) remarked that the goal of meta-analysis is to provide a developmental picture of a particular domain in the most systematic, unbiased manner possible. I suggest that studies with small effect sizes play an important role in pushing researchers to carry out further probing to enhance their interpretations. As shown above, the magnitude of effect sizes was not independent of the ways researchers attempted to assess L2 development. I also suggest that

perspectives from qualitative SA studies can generate new insights into SA learning, not so much in terms of people's movements across physical locations, but as performing identities and (re)negotiating selfhood in relation to a particular context (Morita, 2004; Norton, 2000). Finally, in the field of education, an effect size even as small as 0.1 should be considered important (Glass, McGaw, & Smith, 1981; Oswald & Plonsky, 2010).

Further work is required to do justice to issues of L2 proficiency measurement in the SA research tradition. Kinginger (2009) noted that, under the influence of Utilitarianism, researchers have tended to use language tests as a means of proving linguistic improvement (or deterioration) during SA periods. Nevertheless, L2 proficiency itself is a multidimensional construct, which can be operationalized in many different ways (Fulcher & Davidson, 2007). As Moss (1994) emphasized, a language proficiency test should be designed carefully, with thorough consideration of test validity. Nevertheless, little discussion of L2-measurement methods and their consequences for meta-analysis appears in the current SA scholarship. Given the trend for international journals to encourage authors to report effect sizes (Plonsky & Oswald, 2014), it is suggested that SA researchers carefully consider potential influences on effect sizes before choosing a particular L2-measurement method. In so doing, studies may empirically substantiate or refute previous findings, enriching the depth of discussion on the role of SA in L2 development.

Correspondence should be addressed to **Jin-Suk Yang**, OISE, University of Toronto, 10–229, 252 Bloor Street West, Toronto, Ontario, M5S 1V6; e-mail: jinsuk.yang@mail.utoronto.ca.

Acknowledgements

My deepest appreciation goes to Dr. Alister Cumming, whose graduate course at the Ontario Institute for Studies of the University of Toronto (CTL 3001H, Research Colloquium in Second Language Education) provided me with practical guidance as well as strong motivation for initiating this study. I am grateful for his insightful theoretical feedback and meticulous editorial support in refining the manuscripts. I also thank three anonymous reviewers for their helpful suggestions on earlier versions of this article.

Notes

1 Here I would like to emphasize the epistemological differences inherent in the two research traditions. In quantitative research, the operationalization of constructs is a necessary step in conducting statistical analysis. After

defining the domains in question, researchers investigate the relationships among the variables (Hulstijn, 2010). Researchers with a qualitative orientation, on the other hand, understand language learning as part of larger social processes. That is, emphasis could be put either on linguistic or social aspects of language learning, yet it is difficult to dichotomously code qualitative studies in each category. For this reason, the categories I offer as a result of the initial analysis are meant to serve as a very general indicator of the SA research tradition.

2 VARBURUL, from "*var*i*ab*le r*ule*," is a set of statistical analyses mainly used by sociolinguists and historical linguists to identify linguistic patterns of variation between alternative forms in language use (Mitchell & Myles, 2006).

References

Asterisk (*) indicates the studies that were included in the second phase of meta-analysis

Allen, H. (2010). Language-learning motivation during short-term study abroad: An activity theory perspective. *Foreign Language Annals*, *43*(1), 27–49. http://dx.doi.org/10.1111/j.1944-9720.2010.01058.x

Altbach, P., & Bassett, R. (2004). The brain trade. *Foreign Policy*, *44*(144), 30–31. http://dx.doi.org/10.2307/4152974

Amuzie, G.L., & Winke, P. (2009). Changes in language learning be1iefs as a result of study abroad. *System*, *37*(3), 366–379. http://dx.doi.org/10.1016/j.system.2009.02.011

Barron, A. (2003). *Acquisition in interlanguage pragmatics: Learning how to do things with words in a study abroad context*. Amsterdam, The Netherlands: John Benjamins. http://dx.doi.org/10.1075/pbns.108.

Block, D. (2003). *The social turn in second language acquisition*. Edinburgh, UK: Edinburgh University Press.

Block, D. (2007). *Second language identities*. New York, NY: Continuum.

Brecht, R., Davidson, D., & Ginsberg, R. (1995). Predictors of foreign language gain during study abroad. In B. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 37–66). Amsterdam, The Netherlands: John Benjamins. http://dx.doi.org/10.1075/sibil.9.05bre.

Carroll, J.B. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals*, *1*(2), 131–151. http://dx.doi.org/10.1111/j.1944-9720.1967.tb00127.x

Chang, Y.-F. (2010). "I no say you say is boring": The development of L2 pragmatic competence in L2 apology. *Language Sciences*, *32*(3), 408–424. http://dx.doi.org/10.1016/j.langsci.2009.07.002

Cheng, L., & Fox, J. (2008). Towards a better understanding of academic acculturation: Second language students in Canadian universities. *Canadian*

Modern Language Review, *65*(2), 307–333. http://dx.doi.org/10.3138/cmlr.65.2.307

Churchill, E. (2009). Gender and language learning at home and abroad. *JALT Journal*, *31*, 141–158.

Cohen, A.D., & Shively, R.L. (2007). Acquisition of requests and apologies in Spanish and French: Impact of study abroad and strategy-building intervention. *Modern Language Journal*, *91*(2), 189–212. http://dx.doi.org/10.1111/j.1540-4781.2007.00540.x

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition*, *26*(2), 227–248. http://dx.doi.org/10.1017/S0272263104262040

*Collentine, J.G. (2009). Study abroad research: Findings, implications and future directions. In M.H. Long & C.J. Catherine (Eds.), *The handbook of language teaching* (pp. 218–233). New York: Wiley. http://dx.doi.org/10.1002/9781444315783.ch13.

Davis, J. (2007). Resistance to L2 pragmatics in the Australian ESL context. *Language Learning*, *57*(4), 611–649. http://dx.doi.org/10.1111/j.1467-9922.2007.00430.x

DeKeyser, R. (1991). The semester overseas: What difference does it make? *ADFL Bulletin*, *22*(2), 42–48. http://dx.doi.org/10.1632/adfl.22.2.42

DeKeyser, R. (2010). Monitoring processes in Spanish as a second language during a study abroad program. *Foreign Language Annals*, *43*(1), 80–92. http://dx.doi.org/10.1111/j.1944-9720.2010.01061.x

*Dewey, D. (2008). Japanese vocabulary acquisition by learners in three contexts. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *15*, 127–148.

Dewey, D.P., Bown, J., & Eggett, D. (2012). Japanese language proficiency, social networking, and language use during study abroad: Learners' perspectives. *Canadian Modern Language Review*, *68*(2), 111–137. http://dx.doi.org/10.3138/cmlr.68.2.111

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition*, *26*(2), 249–273. http://dx.doi.org/10.1017/S0272263104262052

Dinsmore, T.H. (2006). Principles, parameters, and SLA: A retrospective meta-analytic investigation into adult L2 learners' access to universal grammar. In J.M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 53–90). Amsterdam, The Netherlands: John Benjamins. http://dx.doi.org/10.1075/lllt.13.07din.

Du, H. (2013). The development of Chinese fluency during study abroad in China. *Modern Language Journal*, *97*(1), 131–143. http://dx.doi.org/10.1111/j.1540-4781.2013.01434.x

Duff, P. (2003). New directions in second language socialization research. *Korean Journal of English Language and Linguistics*, *3*, 309–339.

DuFon, M.A., & Churchill, E. (Eds.). (2006). *Language learners in study abroad contexts.* Clevedon, UK: Multilingual Matters.

Duperron, L., & Overstreet, M. (2009). Preparedness for study abroad: Comparing the linguistic outcomes of a short-term Spanish program by third, fourth and sixth semester L2 learners. *Frontiers*, *18*, 157–179.

Freed, B.F. (1995). What makes us think that students who study abroad become fluent? In B.F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123–148). Amsterdam, The Netherlands: John Benjamins. http://dx.doi.org/10.1075/sibil.9.09fre.

Freed, B.F., Segalowitz, N., & Dewey, D.P. (2004). Contexts of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion program. *Studies in Second Language Acquisition*, *26*(2), 275–301. http://dx.doi.org/10.1017/S0272263104262064

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. New York, NY: Routledge. http://dx.doi.org/10.4324/9780203449066.

Glass, G.V., McGaw, B., & Smith, M.L. (1981). *Meta-analysis in social research*. London, UK: Sage.

Heller, M. (2010). The commodification of language. *Annual Review of Anthropology*, *39*(1), 101–114. http://dx.doi.org/10.1146/annurev.anthro.012809.104951

Hernández, T. (2010). The relationship among motivation, interaction, and the development of second language oral proficiency in a study-abroad context. *Modern Language Journal*, *94*(4), 600–617. http://dx.doi.org/10.1111/j.1540-4781.2010.01053.x

Hulstijn, J.H. (2010). Measuring second language proficiency. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 185–200). Amsterdam, The Netherlands: John Benjamins. http://dx.doi.org/10.1075/lllt.27.11hul.

In'nami, Y., & Koizumi, R. (2010). Database selection guidelines for meta-analysis in applied linguistics. *TESOL Quarterly*, *44*(1), 169–184. http://dx.doi.org/10.5054/tq.2010.215253

Isabelli-García, C. (2010). Acquisition of Spanish gender agreement in two learning contexts: Study abroad and at home. *Foreign Language Annals*, *43*(2), 289–303. http://dx.doi.org/10.1111/j.1944-9720.2010.01079.x

Jackson, J. (2008). *Language, identity, and study abroad: Sociocultural perspectives*. London, UK: Equinox.

Jeon, E.-H., & Kaya, T. (2006). Effects of L2 instruction on interlanguage pragmatic development: A meta-analysis. In J.M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 165–211). Amsterdam, The Netherlands: John Benjamins. http://dx.doi.org/10.1075/lllt.13.10jeo.

Kramsch, C., & Whiteside, A. (2007). Three fundamental concepts in second language acquisition and their relevance in multilingual contexts. *Modern Language Journal*, *91*, 907–922. http://dx.doi.org/10.1111/j.1540-4781.2007.00677.x

Keck, C.M., Iberri-Shea, G., Tracy-Ventura, N., & Wa-Mbaleka, S. (2006). Investigating the empirical link between task-based interaction and

acquisition: A meta-analysis. In J.M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 91–131). Amsterdam, The Netherlands: John Benjamins. http://dx.doi.org/10.1075/lllt.13.08kec.

Kinginger, C. (2009). *Language learning and study abroad: A critical reading of research*. London, UK: Palgrave Macmillan. http://dx.doi.org/10.1057/9780230240766.

Kinginger, C. (2013). Identity and language learning in study abroad. *Foreign Language Annals*, *46*(3), 339–358. http://dx.doi.org/10.1111/flan.12037

Lafford, B.A. (2004). The effect of the context of learning on the use of communication strategies by learners of Spanish as a second language. *Studies in Second Language Acquisition*, *26*(02), 201–225. http://dx.doi.org/10.1017/S0272263104262039

Larsen-Freeman, D. (2000). Second language acquisition and applied linguistics. *Annual Review of Applied Linguistics*, *20*, 165–181. http://dx.doi.org/10.1017/S026719050020010X

Liskin-Gasparro, J. (1998). Linguistic development in an immersion context: How advanced learners of Spanish perceive SLA. *Modern Language Journal*, *82*(2), 159–175. http://dx.doi.org/10.1111/j.1540-4781.1998.tb01189.x

Llanes, À., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, *37*(3), 353–365. http://dx.doi.org/10.1016/j.system.2009.03.001

Marqués-Pascual, L. (2011). Study abroad, previous language experience, and Spanish L2 development. *Foreign Language Annals*, *44*(3), 565–582. http://dx.doi.org/10.1111/j.1944-9720.2011.01148.x

Marriott, H. (1995). The acquisition of politeness patterns by exchange students in Japan. In B. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 197–224). Amsterdam, The Netherlands: John Benjamins. http://dx.doi.org/10.1075/sibil.9.13mar.

Martinsen, R.A., Baker, W., Bown, J., & Johnson, C. (2011). The benefits of living in foreign language housing: The effect of language use and second-language type on oral proficiency gains. *Modern Language Journal*, *95*(2), 274–290. http://dx.doi.org/10.1111/j.1540-4781.2011.01180.x

Masuda, K. (2011). Acquiring interactional competence in a study abroad context: Japanese language learners' use of interactional particle Ne. *Modern Language Journal*, *95*(4), 519–540. http://dx.doi.org/10.1111/j.1540-4781.2011.01256.x

Mitchell, R., & Myles, F. (2006). *Second language learning theories*. London, UK: Hodder Arnold.

Mora, J.C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, *46*(4), 610–641.

Morita, N. (2004). Negotiating participation and identity in second language academic communities. *TESOL Quarterly*, *38*(4), 573–603. http://dx.doi.org/10.2307/3588281

Moss, P.A. (1994). Can there be validity without reliability? *Educational Researcher*, *23*(2), 5–12. http://dx.doi.org/10.3102/0013189X023002005

Norris, J.M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*(3), 417–528. http://dx.doi.org/10.1111/0023-8333.00136

Norris, J.M., & Ortega, L. (2007). The future of research synthesis in applied linguistics: Beyond art or science. *TESOL Quarterly*, *41*(4), 805–815. http://dx.doi.org/10.1002/j.1545-7249.2007.tb00105.x

Norton, B. (2000). *Identity and language learning: Gender, ethnicity and educational change*. Harlow, UK: Longman.

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

Oswald, F.L., & Plonsky, L. (2010). Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics*, *30*, 85–110. http://dx.doi.org/10.1017/S0267190510000115

Paige, R.M., Cohen, A.D., & Shively, R.L. (2004). Assessing the impact of a strategies-based curriculum on language and culture learning abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *10*, 253–276.

Pellegrino Aveni, V. (2005). *Study abroad and second language use: Constructing the self*. Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511620584.

Phillipson, R. (1992). *Linguistic imperialism*. Oxford, UK: Oxford University Press.

Plonsky, L., & Oswald, F. (2014). How big is "Big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. http://dx.doi.org/10.1111/lang.12079

Polanyi, L. (1995). Language learning and living abroad: Stories from the field. In B.F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 271–292). Amsterdam, The Netherlands: John Benjamins. http://dx.doi.org/10.1075/sibil.9.17pol.

Ranta, L., & Meckelborg, A. (2013). How much exposure to English do international graduate students really get? Measuring language use in a naturalistic setting. *Canadian Modern Language Review*, *69*(1), 1–33. http://dx.doi.org/10.3138/cmlr.987

Rees, J., & Klapper, J. (2007). Analyzing and evaluating the linguistic benefit of residence abroad for UK foreign language students. *Assessment and Evaluation in Higher Education*, *32*(3), 331–353. http://dx.doi.org/10.1080/02602930600801860

Regan, V. (1995). The acquisition of sociolinguistic native norms: Effects of a year abroad on second language learners of French. In B. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 245–268). Amsterdam, The Netherlands: John Benjamins. http://dx.doi.org/10.1075/sibil.9.15reg.

Regan, V. (1998). Sociolinguistics and language learning in a study abroad contexts. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *4*, 61–91.

Regan, V., Howard, M., & Lemée, I. (2009). *The acquisition of sociolinguistic competence in a study abroad context*. Bristol, UK: Multilingual Matters.

Rehner, K., & Mougeon, R. (2003). The effect of educational input on the development of sociolinguistic competence by French immersion students: The case of expressions of consequence in spoken French. *Journal of Educational Thought*, *37*(3), 259–281.

Roberts, C., Byram, M., Barro, A., Jordan, S., & Street, B. (2001). *Language learners as ethnographers*. Clevedon, UK: Multilingual Matters.

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J.M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133–164). Amsterdam, The Netherlands: John Benjamins.

Sasaki, M. (2011). Effects of varying lengths of study-abroad experiences on Japanese EFL students' L2 writing ability and motivation: A longitudinal study. *TESOL Quarterly*, *45*(1), 81–105. http://dx.doi.org/10.5054/tq.2011.240861

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. http://dx.doi.org/10.1017/S0272263104262027

Serrano, R., Llanes, A., & Tragant, E. (2011). Analyzing the effect of context of second language learning: Domestic intensive and semi-intensive courses vs. study abroad in Europe. *System*, *39*(2), 133–143. http://dx.doi.org/10.1016/j.system.2011.05.002

Shardakova, M. (2005). Intercultural pragmatics in the speech of American L2 learners of Russian: Apologies offered by Americans in Russian. *Intercultural Pragmatics*, *2*(4), 423–451. http://dx.doi.org/10.1515/iprg.2005.2.4.423

Shedivy, S.L. (2004). Factors that lead some students to continue the study of foreign language past the usual 2 years in high school. *System*, *32*(1), 103–119. http://dx.doi.org/10.1016/j.system.2003.09.008

Siegal, M. (1995). Individual differences and study abroad: Women learning Japanese in Japan. In B.F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 225–244). Amsterdam, The Netherlands: John Benjamins. http://dx.doi.org/10.1075/sibil.9.14sie.

Stewart, J. (2010). Using e-journals to assess students' language awareness and social identity during study abroad. *Foreign Language Annals*, *43*(1), 138–159. http://dx.doi.org/10.1111/j.1944-9720.2010.01064.x

Taguchi, N. (2011). The effect of L2 proficiency and study-abroad experience on pragmatic comprehension. *Language Learning*, *61*(3), 904–939. http://dx.doi.org/10.1111/j.1467-9922.2011.00633.x

Tokowicz, N., Michael, E., & Kroll, J. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition*, *7*(3), 255–272. http://dx.doi.org/10.1017/S1366728904001634

Trenchs-Parera, M. (2009). Effects of formal instruction and a study abroad on the acquisition of native-like oral proficiency. *Canadian Modern Language Review*, *65* (3), 365–393. http://dx.doi.org/10.3138/cmlr.65.3.365

Vatz, K., Tare, M., Jackson, S., & Doughty, J.C. (2013). Aptitude-treatment interaction studies in second language acquisition: Findings and methodology. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 273–292). Amsterdam, The Netherlands: John Benjamins. http://dx.doi.org/10.1075/lllt.35.11vat.

Wang, C. (2010). Toward a second language socialization perspective: Issues in study abroad research. *Foreign Language Annals*, *43*(1), 50–63. http://dx.doi.org/10.1111/j.1944-9720.2010.01059.x

Wilkinson, S. (1998). Study abroad from the participants' perspective: A challenge to common beliefs. *Foreign Language Annals*, *31*(1), 23–39. http://dx.doi.org/10.1111/j.1944-9720.1998.tb01330.x

Yang, J.-S., & Kim, T.-Y. (2011). Sociocultural analysis of second language learner beliefs: A qualitative case study of two study-abroad ESL learners. *System*, *39* (3), 325–334. http://dx.doi.org/10.1016/j.system.2011.07.005

Yashima, T., Zenuk-Nishide, L., & Shimizu, K. (2004). The influence of attitudes and affect on willingness to communicate and second language communication. *Language Learning*, *54*(1), 119–152. http://dx.doi.org/10.1111/j.1467-9922.2004.00250.x

Appendix A

Sample of Coding Scheme

ID No. _____

1. Publication Characteristics

   Author(s): _____

   Title: _____

   Journal: _____

   Edited volume: Year _____ volume _____ Pages_____

   Source of the reference

   ERIC _____ EBSCOhost _____ ProQuest _____

   Manual Search _____

2. Participant Characteristics

   In-group comparison _____ Between-group comparison _____

   Qualitative research_____

   Control Group (or target groups)

   Mean age _____ and *SD* _____

   Adolescent _____ Adult _____ Mixed _____

   Gender F_____ M _____ Mixed _____

   Experimental Group

   Mean age _____ and *SD* _____

   Adolescent _____ Adult _____ Mixed _____

   Gender F_____ M _____ Mixed _____

   Native Language (Control Group)

   English _____ Chinese_____ Korean _____Spanish _____

   Japanese _____ Others_____

Native Language (Experimental Group)

English _____ French_____ Korean _____

Spanish _____ Japanese _____ Others _____

Target Language

English _____ French_____ Korean _____

Spanish _____ Japanese _____ Others _____

Length of Time in Target Language Country

_____years _____months

3. Design Characteristics

Sample size

Total _____

Type of Study abroad _____

Language structure targeted

Vocabulary_____Grammatical _____Sociolinguistic _____

Pragmatic _____Strategic _____

Measurement _____

Results comparison

Control vs comparison

mean_____*SD* ____ vs Mean_____ *SD* ____

4. Statistical Information

Type _____

Control group

no of participants _____means_____ *SD*_____

Experimental group

no of participants_____ means_____ *SD*_____

Appendix B

Distribution of first language (nationality) of SA studies

Nationality	Number of Participants
Arabic	1
Chinese	10
English[4]	30
Hmong	1
Russian	1
Japanese	6
Korean	8
French	2
Spanish	7
German	2
Swedish	2
Samoan	2
Somalian	2
Taiwanese	2
Iranian	1
Indonesian	1
Philippines	1
Romanian	1
Thai	2
Ethiopian	1
Malaysian	1
Bosnian	1
Hmong	1
Unknown	4

Note: In some studies (e.g., Shedivy, 2004), researchers did not specify the nationality to which students belonged when providing basic information on the participants' first language(s). This is conspicuous when the participants' first language is English. In such cases, I coded the students according to their first language.

Appendix C

Features of 11 included studies

	Author/ Publication Year	Study Focus	Language First-Target	Assessment Procedure/ Instrument
1	Collentine (2004)	Morphosyntactic & lexical proficiency	English-Spanish	Oral proficiency test/ corpus analysis
2	Dewey (2008)	Vocabulary proficiency	English-Japanese	Vocabulary matching test/ Situational vocabulary test
3	Isabelli-García (2010)	Grammatical knowledge	English-Spanish	Stimulated oral proficiency interview
4	Lafford (2004)	Communication strategies	English-Spanish	Oral proficiency interview
5	Marqués-Pascual (2011)	Morphosyntactic development	English-Spanish	Oral narratives
6	Martinsen, Baker, Bown & Johnson (2011)	Oral proficiency	Not specified-French, German, Russian, Japanese	Oral proficiency interview
7	O'Brien, Segalowitz, Freed, & Collentine (2007)	Phonological memory	English-Spanish	SNWR (phonological memory) Oral proficiency interview
8	Sasaki (2011)	Writing proficiency/ motivation	Japanese-English	Argumentative composition/English composition profile
9	Segalowitz & Freed (2004)	Oral proficiency	English-Spanish	Oral proficiency interview
10	Serrano, Llanes, & Tragant (2011)	Writing & Oral proficiency	Spanish-English	Composition/Oral narrative
11	Taguchi (2011)	Pragmatic proficiency	Japanese-English	Pragmatic listening test