

1. Introduction

Study abroad (SA) describes the act of travelling to a foreign country for the purpose of study. In common usage the term encompasses both cultural exchange programmes (e.g. U.S. students studying in the UK) and programmes designed to promote language acquisition, such as tertiary-level academic exchanges for university language majors (e.g. the ERASMUS programme in Europe). Academic SA (i.e. programmes delivered by a university/FE college) can last from one semester to one or two academic years. Further variation exists in non-academic SA for language acquisition (i.e. programmes delivered by private language institutions): programmes include work internships, summer language courses, intensive and extensive language courses, language holiday homestays (whereby the student lives with a language tutor abroad and receives one-to-one tuition in the home), and educational trips abroad. Non-academic SA ranges from a long weekend to over a year spent living and studying in the target language community.

SA is unique as a language learning context because of the constant availability of L2 contact. The majority of a SA learner's time is spent outside of the classroom and students often cohabit with people who do not speak their L1 (e.g. with a host family or in mixed-L1 halls of residence). Indeed, SA courses at private language institutions are sometimes sold on the promise of out-of-class language contact and resulting native-like L2 mastery. For example, Eurolingua (2015) guarantee native-like speaking proficiency as a result of studying abroad on their programmes: "No matter what your level, you will soon be speaking fluently, with understanding, ease and confidence. By taking part in one of our study abroad programs...you will be armed with an authentic accent." Misleading (and often inaccurately worded) promises are also made about the L2 contact that learners will experience on SA: "Most people choose study English in an English-speaking country because it is the best way to improve in a short time their English skills, you will be able to practice your English everywhere not only in classes" (LearnEnglishGuide, 2015).

However, heightened exposure to the L2 does not always materialise (Pellegrino, 1998; Wilkinson, 1998): Learners who live alone or with speakers of their L1 may sustain little L2 contact outside of class, and those living with native English speakers may not interact with their cohabitants. Individual differences (e.g. motivation, personality, age) may also withhold some learners from the rich and frequent exposure to the L2 that SA promises (Tanaka, 2004; Woodman, 1998). Furthermore, there is a dearth of research into the specific types of out-of-class language contact that SA learners encounter, and neither has the relationship between specific types of out-of-class contact and specific linguistic outcomes been comprehensively addressed. Are some types of language contact more beneficial to vocabulary acquisition than others to engage in? Which are likely to beget meaningful language processing? It is questions such as these that remain largely unanswered yet have the potential to dramatically impact upon the linguistic outcomes of SA. Therefore, this study set out to determine the specific types of out-of-class language contact that SA learners in England identify with, and whether a relationship exists between that contact and vocabulary gain. Moreover, as one could hypothesize that the location and length of a sojourn abroad may impact upon out-of-class L2 contact, so the role of these

variables in the relationship between out-of-class contact and vocabulary gain was explored.

2. Literature Review

2.1 Study abroad and vocabulary gain

The literature on the impact of SA on second language acquisition has suggested that SA learners are significantly more orally fluent after SA in comparison with learners who studied in the home country context (DeKeyser, 1991; Freed, 1995; Lafford, 2004; Segalowitz & Freed, 2004). Moreover, L2 learners experience a significant impact of SA on the development of their vocabulary knowledge (Milton & Meara, 1995; DeKeyser, 1991; Ife *et al.*, 2000; Dewey, 2008; Llanes & Muñoz, 2009). However, for other components of second language competence (e.g. listening comprehension, grammatical control) the benefit of SA over 'at home' settings has not yet been firmly established (Collentine, 2004; DeKeyser, 1991; Freed *et al.*, 2003; Díaz-Campos, 2004; Dewey, 2004; Mora, 2008).

The present study focuses on the impact of SA on gains in vocabulary knowledge. Research has provided strong evidence as to the primacy of vocabulary knowledge in L2 competence. Vocabulary knowledge has been found to predict reading comprehension (Laufer, 1992; Hsueh-Chao & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010) and listening comprehension (Stæhr, 2009; van Zeeland & Schmitt, 2013). There is also robust evidence in favour of the positive impact of vocabulary knowledge on productive ability in the L2 (Laufer & Nation, 1995; Stæhr, 2008; Koizumi & In'nami, 2013). Research suggests that over and above its relationship to the four skills, breadth of vocabulary knowledge may impact on grammatical competence (Myles, Hooper & Mitchell, 1998; Bardovi-Harlig, 2002) and phonological control (Bundgaard-Nielsen, Best & Tyler, 2011a, 2011b).

A large number of studies on vocabulary knowledge and SA have found evidence of gains. Studies have either taken a prospective, within-subjects design or a between-groups design that compares the gains of SA learners with participants in other settings. In terms of within-groups studies, SA has been found to have a positive impact on rate of lexical acquisition (Milton & Meara, 1995), lexical organization and vocabulary size (Ife, Vives Boix & Meara, 2000), and development of receptive vocabulary knowledge (Pizziconi, 2013). However, Fitzpatrick's (2012) study of a L1 Chinese SA learner of L2 English indicated that the relationship between SA and lexical development may not be entirely straightforward: her participant's acquisition of individual lexical items was subject to frequent, non-linear change.

With regard to between-groups designs, there is a convincing body of evidence to suggest that SA is more beneficial to L2 lexical development than at home study (Laufer & Paribakht, 1998; Dewey, 2008; Foster, 2009; Serrano, Llanes & Tragant, 2011). However, the evidence is not entirely clear cut: Dewey (2008), in his study of at home, domestic immersion and SA learners of L2 Japanese found no significant difference between the lexical outcomes of the domestic immersion and SA participants. Similarly, Serrano, Llanes and Tragant's (2011) study of L1 Spanish learners of English found no significant

differences between the gains of the SA group and an at home group who had studied the L2 intensively for the same period of time. Furthermore, Collentine's (2004) study of n=46 at home and SA learners of Spanish failed to find any significant differences between the groups in terms of frequency of use in oral production of conjunctions, pronouns, prepositions, adverbs, verbs and nouns.

2.2 Out-of-class language contact

Many language learners view a stay in the target language community as the *sine qua non* of second language mastery and believe that heightened exposure to the L2 outside of the classroom is what makes the difference. Out-of-class language contact has been measured most frequently in the SA literature using a questionnaire called the Language Contact Profile, henceforth the LCP. The LCP was developed by Freed, Dewey, Segalowitz and Halter (2004) for use with L1 English learners of L2 Spanish and comprises two sections; one for use before the SA and the other to be administered after the SA has ended. The pre-SA section collects demographic data and information about the respondent's proficiency in the L2. There then follows 13 items that describe contact with the L2 in the home country to which the respondent chooses from a 5-point frequency scale (from 'almost never' to 'daily'). The post-SA section collects data about the respondents' living arrangements during SA and then presents 41 items that describe out-of-class language contact scenarios (e.g. I read magazines and newspapers in Spanish / I speak to service personnel in Spanish). The respondent must state on average how many days per week during the SA they engaged in each scenario, and how many hours per day (from zero to five+). The items are grouped according to skill; in other words, all of the scenarios that involve speaking are listed together, followed by reading, listening and writing.

For such a widely used instrument there is little published evidence of the validity and reliability of the LCP. Martinsen (2011) adapted the LCP for use in a study of the impact of SA on cultural sensitivity and reported a high Cronbach's alpha coefficient of .83, yet failed to include any specific details regarding how the instrument had been adapted. Author (2015) used an adaptation of the LCP in a study of n=95 SA learners in the UK who comprised a range of L1s and nationalities. She adapted the response scale to a 5-point 'how true of me' scale in order to avoid collecting unreliable frequency data: Dörnyei (2003) states that frequency scales of items of a different nature should not be treated cumulatively. Moreover, the wording of questionnaire items in the original instrument that began with "I try to..." were amended so that the items represented scenarios encountered 'naturally' by learners as opposed to strategies they employ to increase their out-of-class language contact. Similarly to Martinsen's (2011), Author's (2015) version of the LCP yielded a high Cronbach alpha coefficient (0.85).

The LCP and similar instruments have been used to provide evidence of a positive relationship between language contact and a number of aspects of linguistic competence: oral fluency (Yager, 1998; Freed, Segalowitz & Dewey, 2004; Segalowitz & Freed, 2004; Hernández, 2010), lexical knowledge (Segalowitz & Freed, 2004; Dewey, 2008), phonological control (Muñoz & Llanes, 2014), intercultural competence (Martinsen, 2011). The findings of these studies tend to suggest that the more contact with the L2 outside of class, the better, and

that engagement with different types of out-of-class contact impacts upon different elements of L2 competence. For example, the proportion of time spent speaking with native L2 speakers has been found positively to correlate with gains in degree of foreign accent (Muñoz & Llanes, 2014) and intercultural sensitivity (Martinsen, 2011).

To ascertain whether a correlation existed between language contact and gains in vocabulary knowledge, Dewey (2008) administered the LCP (Freed *et al.*, 2004) to n=56 L2 Japanese learners in three settings (EFL, domestic immersion and SA). He administered three vocabulary measures: a picture matching test, a situational vocabulary test and an adaptation of the Vocabulary Knowledge Scale (Paribakht & Wesche, 1993). There was a significant positive correlation between time spent speaking Japanese (with friends or a host family) and gains in situational vocabulary knowledge. Moreover, a significant negative correlation was found between reading on the Internet in the L2 (e.g. surfing the web; reading emails) and gains in depth of vocabulary knowledge. Conversely, Segalowitz and Freed (2004) studied n=47 L1 English L2 Spanish learners in 'at home' and SA contexts over one academic term to find no significant relationship between speed and accuracy of lexical access and out-of-class language contact.

The literature suggests that SA learners experience less interactive out-of-class contact with native speakers than they might have expected (DeKeyser, 1986; Polanyi, 1995; Rivers, 1998; Wilkinson, 1998). The variables that can inhibit or modify a SA learner's contact with the L2 include gender, personality, motivation, social distance, proficiency and age. Living arrangements have also been shown to play a key role: For example, as stated by Author (2015), language learners who live by themselves during the SA, or who live with speakers of their native tongue, may choose to navigate their time outside of the classroom predominantly in the L1. Moreover, even where a SA learner lives in a L2 setting (such as with a host family) they may choose, as found by Kinginger (2008), to disengage from the L2 as a result of negative interactional experiences with their cohabitants or with others. Similarly, it is conceivable that a lack of interaction between the host family and the SA learner could stem from disinterest or emotional distancing on the part of the host family. Certainly, the promise of supportive out-of-class L2 interaction suggested by a stay with a host family has been shown not always to materialise (Wilkinson, 1998; Twombly, 1995). Pellegrino Aveni's (2007) exploration of the experiences of North American SA learners in Russia highlighted that a learner's sense of self can be inhibited by the culture they meet and/or their ability to use the L2 to express that sense of self.

2.3 Location and study abroad

Location has received relatively little attention in the SA literature and to date there are no studies which have investigated the relationship between location of SA experience and vocabulary gain. Dewey, Bown, Baker, Martinsen, Gold and Eggett (2014) state that the vast majority of SA research has sampled participants who studied abroad in the same location. Moreover, where a sample has been divided by location, location has generally not been treated as a variable (e.g. Magnan & Back, 2007; Dewey, Belnap & Hillstrom, 2013). Dewey *et al.*'s (2014) study is one exception because their sample compared the out-of-

class language contact of SA learners in Spain, Mexico, France, Russia, China and Egypt. Regression analysis revealed that studying in Mexico and Egypt was predictive of out-of-class contact [$R^2=.103$, $F(2,113)=6.40$, $p=.002$]. However, both of these SA programmes provided organised opportunities for out-of-class L2 contact and thus, it was not location *per se* that impacted out-of-class contact, but rather features of the SA programmes in those particular locations. Dewey, Ring, Gardner and Belnap (2013) studied $n=71$ SA learners of L2 Arabic in Jordan and Egypt to find that location was a top influencing factor on the formation of social networks: living in close proximity to native L2 speakers of the same age group yielded broader and deeper social relationships for the participants in Jordan. Watson, Siska and Wolfel (2013) found that only their participants who studied abroad in China and Western Europe (as compared with participants in Eastern Europe, Latin America and the Middle East) made significant gains in terms of regional knowledge. However, as there were no controls for living arrangements, programme features or language contact, it is not clear whether this difference was in fact due to location.

2.4 Length of stay and study abroad

Given the relative affordability and convenience of a shorter SA experience in comparison with a longer stay, this variable has received quite considerable attention in the literature. However, the synthesised findings are so far largely inconclusive, and varying lengths of stay have been found to have varying degrees of impact on different components of linguistic proficiency. Some studies have measured the impact of length of stay on a range of linguistic outcomes. For example, over a period of fifteen years Davidson (2010) conducted a study of $n=1,881$ U.S. learners of Russian in Russia for either two, four or nine months. He administered measures of grammar knowledge, oral proficiency, reading, listening and writing and set out to investigate the impact on these outcomes of length of stay, gender, age and initial proficiency. Davidson (2010:17) states that his study found 'clear relationships' between the independent variables of length of stay and initial proficiency on the outcome measures, but this relationship appears to be drawn from the larger percentage gain of the long stay participants than the short or medium stay group. However, the author does not report *t*-tests with reference to the pre and post scores, so we cannot rely on their being statistically significantly different. Moreover, Davidson (2010) does not include length of stay as a predictor variable in his regression model, but rather conducts one regression for each of the three length of stay groups. As such, there is no convincing evidence to support his claim that length of stay is related to linguistic gain.

In another study of the impact of length of stay on multiple linguistic outcomes, Baró and Serrano (2011) compared the oral and written development of $n=46$ L1 Spanish undergraduate learners of L2 English, who travelled to the UK for academic SA of either two or three months. Pre and post-SA measures of oral and written fluency were administered within a week of departure to the UK and within a week of return to Spain. They comprised a 15-minute written composition and an oral picture narration task. Analysis of the outcome texts focused on fluency, complexity and accuracy. There were no significant differences between the groups on written or oral development. The authors

suggested therefore that a difference of only one month is insufficient time for any advantage of length of stay to have a detectable impact.

Other studies have investigated the role of length of stay in the development of specific components of second language acquisition. For example, Mora and Valls-Ferrer (2012) investigated the relative development of oral fluency, accuracy and complexity over a three-month SA to find that only gains in fluency were significantly different during this time frame. Højen (2003) studied the development of participants' perceived foreign accent to find that the longer the stay, the less accented the L2 oral production. A similar pattern was uncovered by Sasaki (2009), who found that gains in the L2 writing proficiency of her L1 Japanese sample increased by length of stay. More evidence in support of 'the longer, the better' was found by Félix-Brasdefer (2004) in a study of the impact of SA on the pragmatic development of four length of stay sample groupings.

There is evidence of an impact of length of stay on vocabulary development: Iñe, Vives Boix and Meara (2000), in their longitudinal study of the impact of SA and proficiency level on lexical development, measured the type and rate of L2 Spanish vocabulary acquired by $n=36$ native English speakers, who were studying Spanish at a university in Spain over a period either of one or two semesters. They found that length of stay was positively correlated with gains in lexical organisation as measured by a word association test ($r=.5471$, $p<.001$); however, as only $n=11$ participants stayed for two semesters (i.e. comprised the long stay group), this evidence should be considered as tentative.

2.5 Summary of literature review

SA has been shown to positively impact on a number of facets of vocabulary knowledge and for the most part to be more beneficial to vocabulary knowledge gain than domestic learning contexts. There is evidence to suggest, however, that domestic L2 immersion is similarly beneficial. Whilst evidence exists in support of the positive impact of out-of-class language contact during SA on linguistic outcomes, a firm relationship between out-of-class contact and vocabulary knowledge gain has yet to be established. Moreover, it may be that not all learners are offered the same out-of-class contact opportunities, and learners may choose not to take advantage of the opportunities that do present themselves. The little evidence so far found with regard to location of SA and linguistic gain is either ungeneralisable or has been confounded by programme variables. Regarding length of stay, in general terms the evidence points to a benefit of a longer stay in the target language community yet there is only tentative evidence of this impact on vocabulary knowledge gain.

In light of the current evidence on the relationship between out-of-class language contact and vocabulary knowledge gain in a SA context, the present study was guided by the following research questions:

1. What types of out-of-class language contact do study abroad learners identify with most?
2. Is there a relationship between different types of out-of-class language contact and vocabulary knowledge gain in a study abroad context?

2a. Do location and/or length of stay play a role in this relationship?

3. Methodology

3.1 Sample

The sample (n=241) comprised adult (18-53 years old) SA learners of L2 English (n=98 males and n=143 females) who were studying at private language institutions (rather than taking part in an academic exchange such as ERASMUS) in either Oxford (n=112) or London (n=129). By choosing two cities of very different sizes and cultural makeups, the researcher hoped to ascertain whether location has an impact upon out-of-class L2 contact. E.g. in a larger, more multicultural city such as London, a larger number of SA learners may reside within their L1 community and/or interact more with L2 speakers of their L1, and thus may have a different experience of out-of-class language contact than learners elsewhere.

The sample was also grouped by length of stay. To participate, a learner was required to be undertaking at least six weeks of SA in order to allow sufficient time for them to experience the out-of-class L2 contact scenarios which were described in the questionnaire items (listed in table 1). A previous study of a very similar population (Author, 2015) suggested that the average non-academic SA is less than 20 weeks in length; therefore the longest period of participation in the present study was 20 weeks. N=55 participants constituted the short stay group (6-10 weeks), n=56 the medium stay group (11-15 weeks) and n=130 the long stay group (16-20 weeks).

The sample included learners from China, Japan, South Korea, Taiwan, Thailand, Saudi Arabia, Turkey, Brazil, Colombia, Argentina, France, Spain, Germany, Switzerland, Cyprus, the Czech Republic and Russia. There was no restriction on nationality or L1 in order that the out-of-class L2 contact of non-academic SA learners in the UK as a whole group (i.e. the population being studied) might be identified. The participants were predominantly in their 20s and 30s, and their reason for coming to England to study English was to increase their job prospects in their home country, for the enjoyment of learning English as a L2, and/or for experience of living abroad. Participants were receiving 15-25 hours per week of general English instruction (i.e. preparation for everyday situations and language in English) and therefore, if a 14-hour waking day is presupposed, the vast majority of their time was spent outside of the classroom.

A participant's proficiency level (measured by the outcome of the placement test administered by their language institution on arrival to the UK) was a minimum of the Common European Framework of Reference (CEFR) level B1 (low-intermediate) to ensure they had the linguistic ability to complete the questionnaire. The proficiency of the sample at the point of recruitment to the study ranged from CEFR level B1 (low intermediate) to C2 (mastery). The majority of the participants tested at level B2 (high intermediate).

All n=241 participants completed the time 1 vocabulary test and the questionnaire but, due to participant mortality, only n=196 completed the time 2 vocabulary test (c.f. 4.3).

3.2 Instruments

3.2.1 Vocabulary test

The vocabulary measure was adapted from Nation's (1990) Vocabulary Levels Test, and Laufer and Nation's (1999) productive version of the Vocabulary Levels Test. Nation's (1990) levels test is a test of receptive vocabulary breadth developed on the premise that learners are likely to know more high-frequency words than low-frequency words. The receptive test comprises 36 words and 18 definitions at each frequency level. Each question presents the test taker with six lemmas and three definitions, so that three of the lemmas function as distractors. As such, it is a test of passive recognition of the orthographic form-meaning link and of knowledge of meaning in terms of concepts, referents and associations. Laufer and Nation's (1999) productive levels test comprises 18 words at each frequency level. Each question is a sentence in which the target word is incomplete; respondents must complete the word. The productive test is a measure of active word knowledge in terms of meaning, appropriateness, collocations, grammatical position and orthographic form (Fitzpatrick & Meara, 2004). The levels tests are widely used in both testing and research contexts across the globe (Schmitt, Schmitt & Clapham, 2001) and as such a number of studies have sought to determine their validity and reliability and have found favourable evidence in these respects (Read, 1988; Beglar & Hunt, 1999; Laufer & Nation, 1999; Schmitt, Schmitt & Clapham, 2001; Xing & Fulcher, 2007; Platzer, 2012; Abdullah, Puteh, Azizan, Hamdan & Saude, 2013).

The levels tests were chosen as the vocabulary measure in this study for a number of reasons. Firstly, combined they cover receptive and productive vocabulary, so provide the means to test improvement in both forms of knowledge. Secondly, because in the receptive test the six words in each question are semantically very different, even partial word knowledge is sufficient to answer the question correctly (Beglar, 2000). This sensitivity complements well the out-of-class vocabulary acquisition which is approached in this study; outside of the classroom setting it is far less likely that a learner will fully understand a word/phrase because they may not be exposed to comprehensive or accurate definitions, there is unlikely to be any checking of understanding, and the opportunity to use the word/phrase may never occur. Similarly in the productive test, the measure is not entirely of productive knowledge because in giving the first letters of the target word, receptive knowledge is employed to recognise that word-part. Therefore, there is a mirror between the two tests in terms of sensitivity to knowledge. Thirdly, unlike other widely-used measures of testing lexical breadth, such as the Eurocentres Vocabulary Size Test (Meara & Jones, 1990), the levels tests do not rely on a learner's ability to evaluate their own knowledge of a particular word. This means that the quantitative data generated by the levels test is a snapshot of actual knowledge rather than perceived knowledge. Finally, the levels test is an L2-only test that is a necessary feature in a study that employs a sample of mixed L1s.

To reduce the time it would take to complete the vocabulary measure in this study, the eighteen target items per level in the productive test were decreased to ten items per level. Some of the eight items omitted at each level

were chosen because they were specifically American English words (e.g. 'diaper'), which were unlikely to be acquired in out-of-class contact with British English. The others were selected at random. The test was piloted with n=8 study abroad learners who matched the population of this study and the results of the pilot showed that the test was sufficiently sensitive to discriminate between SA learners whose receptive and productive vocabulary size was different from one another. Moreover, the piloting demonstrated that there was no ceiling effect of the test. In both parts of the test (receptive and productive), instead of presenting the items in order of word level the order was randomly mixed so that participants were less likely to give up or not attempt all of the questions once they had reached a difficult stage. However, there was a preponderance of items from the 2000 and 3000 word frequency levels at the beginning of each part to prevent participants from becoming disheartened early on.

3.2.2 Out-of-class language contact questionnaire

The questionnaire used to measure out-of-class language contact in the present study was an adaptation of the LCP (Freed *et al.*, 2004), which is described in detail in Author (2015). The key adaptation was a modification of the response scale from a measure of frequency to a 5-point 'how true of me' rating scale that ranged from 'This is very true of me' to 'This is not at all true of me'. Thus, the adapted questionnaire that was used in this study measured a participant's level of identification with the type of out-of-class L2 contact described in each questionnaire item. This adaptation was made in light of Dörnyei's (2003) statement that frequency scales of items of a different nature should not be treated cumulatively.

3.3 Procedure

3.3.1 Recruitment

Four private language institutions in London and three in Oxford agreed to participate in the study. Newly arrived language learners who matched the sampling frame of the study were invited to attend a recruitment meeting at their institution, led by the researcher. Those who agreed to participate were formally recruited to the study and completed the time 1 vocabulary test within two weeks of their arrival to the UK.

3.3.2 Vocabulary testing

The time 1 test was administered on site at a participant's language institution, in a classroom and under test conditions. Participants were told that they would take another vocabulary test at the end of their participation, but (to negate any practice effect) not that the second test would contain the same items. Participants were advised that there was only one correct answer for each question. For part 2 (productive knowledge), they were told not to be overly concerned with spelling; if the researcher was able to understand the word and it was the correct answer, they would earn a mark. The participants were given

40 minutes in total to complete both parts of the test and were independent in their decision about how much time within those 40 minutes to spend on each part, and on the test in total. The administration of the time 2 test matched exactly that of time 1, and took place no more than two weeks prior to the end of a participant's participation in the study. The receptive vocabulary tests were marked solely by the researcher as there was only one possible correct match for each lexical item tested. The productive tests were double-blind marked by the researcher and an inter-rater.

3.3.3 Questionnaire administration

Participants completed the questionnaire, which measured their identification with different out-of-class language contact scenarios, no sooner than two weeks after their arrival in the UK and no later than two weeks before the end of their participation in the study. The questionnaire was administered in their language institutions and in the presence of the researcher. A period of at least 2 weeks' residence in the target language community before completion of the questionnaire was decided upon because a fortnight's stay allowed enough time to have passed for participants to experience the scenarios about which they were questioned in the instrument but was a short enough period that the short stay group (6-10 weeks) could be included. Respondents were encouraged to ask questions related to the wording of the items but the researcher did not influence the content of their responses in any way.

4. Findings

4.1 What types of out-of-class language contact do study abroad learners identify with most?

The internal consistency of the data from the questionnaire, which comprised 28 items pertaining to out-of-class language contact, was calculated using Cronbach's alpha, finding a high coefficient of .816. The mean scores of the items were calculated to determine the order in which the scenarios were most identified with by the sample (illustrated in table 1)¹ and subsequently exploratory factor analysis was applied to determine whether the scores would reveal any underlying structure. The correlation matrix revealed numerous correlation coefficients of .3 and above. The Kaiser-Meyer-Olkin measure of sampling adequacy was .576 and Bartlett's (1954) Test of Sphericity was significant ($p=.000$), supporting the factorability of the matrix. Nine factors recorded eigenvalues above 1 and explained a total of 78.4% of the variance. A scree plot was mapped and the factor matrix inspected to select which factors to retain. The scree plot revealed a sharp elbow in the plot after factor 4 and reference to the factor matrix revealed that only the first three factors comprised a majority of items that loaded onto them at more than .4. Therefore, the first three factors were retained, and a Varimax rotation conducted better to interpret them.

¹ Table 1

Nine of ten of the scenarios that loaded onto factor 1 involved speaking and/or listening to spoken language (e.g. I listen to English songs outside of class / I speak English to the people I live with). Therefore, factor 1 was labelled 'Interactive'. The majority of the eight scenarios loading onto factor 2 illustrated an application of literacy skills in English (e.g. I read timetables, announcements, posters, menus etc. in English outside of class / Overall, I think I do a lot of writing in English outside of class): all bar two involved reading or writing. The scenarios that did not involve reading or writing neither involved the use of English in that they were conducted in the L1 or a L3 (e.g. I speak my native language to native or fluent speakers of my native language). For this reason, factor 2 was labelled 'Literate'. The six scenarios which appeared in factor 3 all related to the production or comprehension of narratives in some way: they each involved either reception of narratives (through watching movies and using subtitles to enhance comprehension of the narrative) or represented situations in which a learner might narrate their experiences to others (e.g. I read subtitles while I am watching TV or movies in English outside of class / I write emails in English outside of class). As such, factor 3 was labelled 'Narrative'.

There was no significant difference in overall identification with out-of-class language contact between participants in Oxford (M=1.62, SD=.12) and participants in London (M=1.59, SD=.17; $t(238)=1.40$, $p=.162$). However, in response to 11 of the 28 scenarios, the Oxford and London groups' means were different. A Bonferroni adjustment found that three of the 11 differences reached significance, achieved at the $p<.002$ level². Why might study abroad learners in Oxford identify more highly with scenarios involving watching television, reading subtitles and writing in the L2 than their counterparts in the capital? In investigating this question a chi-square test of independence revealed that the proportion of participants that lived with people who did not speak their L1 (i.e. with a host family or in a mixed-nationality house-share) was significantly larger among participants in Oxford than participants in London [$\chi^2(2, N=241)=6.414$, $p=.04$]. By extension, therefore, those participants whose cohabitants did not share their L1 identified more highly with scenarios that involve watching television and subtitles and writing in the L2.

A one-way between-groups ANOVA revealed a significant difference in identification with out-of-class contact for the three length of stay groups [$F(2, 237)=12.06$, $p=.000$, $\eta^2=.09$]. Post-hoc comparisons using the Tukey HSD test indicated that identification with out-of-class contact for the short stay group (M=1.53, SD=.18) was significantly different ($p=.013$) from that of the medium stay group (M=1.62, SD=.10) and the long stay group (M=1.64, SD=.14). The medium and long stay groups did not significantly differ from one another. According to these findings, the short stay group identified more strongly with out-of-class language contact than the medium or long stay group (because in the questionnaire 1=This is very true of me and 5=This is not at all true of me).

4.2 Is there a relationship between different types of out-of-class language contact and vocabulary knowledge gain in a study abroad context? Do location and/or length of stay play a role in this relationship?

² Table 2

The two raters had a 98.5% agreement rate on the productive vocabulary tests, which yielded a Cohen's Kappa value of .97 (SE=.005). The sample as a whole made significant gains on receptive [$t(196)=-5.66, p=.000, \eta^2=.014$], productive [$t(196)=-6.14, p=.000, \eta^2=.016$] and overall (i.e. receptive and productive scores combined) vocabulary knowledge [$t(196)=-7.30, p=.000, \eta^2=.021$].

Regarding location and vocabulary gain, no significant differences were found between participants in Oxford and London. A one-way between-groups ANOVA revealed a significant difference in receptive vocabulary gain for the sample as grouped by length of stay [$F(2, 193)=3.05, p=.050, \eta^2=.03$]. A Tukey HSD test indicated that the mean gain of the medium stay group ($M=4.89, SD=28.99$) was significantly different ($p=.050$) from that of the long stay group ($M=19.08, SD=33.79$). The short stay group ($M=18.71, SD=40.09$) did not differ significantly from either the medium or the long stay group. Similarly to receptive gain, there was a significant difference in overall vocabulary gain for the three groups [$F(2, 190)=3.749, p=.025, \eta^2=.04$]: the mean overall gain of the medium stay group ($M=5.22, SD=21.85$) was significantly different ($p=.019$) from the long stay group ($M=18.04, SD=27.73$). The short stay group ($M=13.18, SD=28.29$) did not differ significantly from either other group. There was no significant difference in productive vocabulary gain between the sample as grouped by length of stay.

In order to investigate further the impact of length of stay on vocabulary gain a repeated measures *t*-test was conducted, which compared the mean scores from the time 1 and time 2 vocabulary tests for each of the three length of stay groups independently. In terms of receptive vocabulary, only the short stay group ($p=.005$) and long stay group ($p=.000$) made a statistically significant gain. The effect size was the same for both groups and was small ($\eta^2=.16$). With regard to productive vocabulary, the medium stay group made a significant gain ($p=.002$) with a small effect size ($\eta^2=.18$), as did the long stay group ($p=.000$) with a small effect size ($\eta^2=.21$), but the short stay group did not make a significant gain ($p=.079$). In terms of overall vocabulary gain, only the short stay and the long stay groups made a significant improvement from time 1 to time 2. For the short stay group ($p=.002$) the effect size of the overall gain was small ($\eta^2=.18$), yet for the long stay group ($p=.000$) the effect size for overall vocabulary gain was large ($\eta^2=.76$). To summarise, there was a statistically significant difference between the sample as grouped by length of stay on their receptive and overall vocabulary gain. The effect sizes in terms of receptive gain were small for both groups that made a significant gain (short and long stay). The effect size for the overall vocabulary gain of the short stay group was small but the effect of SA on the overall vocabulary gain of the long stay group was large.

Correlation analyses of out-of-class language contact (overall and by factor) and vocabulary gain found no statistically significant relationship between the two. Hierarchical multiple regression analysis was used to determine the ability of the three independent variables (out-of-class contact, location and length of stay) to predict vocabulary gain. The three out-of-class contact factor variables were entered as block 1, and location and length of stay were entered together as block 2. Analyses took receptive, productive and overall vocabulary gain as the dependent variable respectively. For receptive

gain, neither model 1 nor model 2 reached significance. Length of stay was the only predictor variable to make a significant contribution to the final receptive model ($\beta = -.182, p = .03$): membership of the medium stay group (11-15 weeks) predicted lower gain. For productive gain, neither model was significant and none of the predictor variables made a significant contribution. For overall vocabulary gain, neither model was significant but length of stay once more made a significant contribution: belonging to the medium stay group predicted lower overall gain ($\beta = -2.49, p = .014$).

5. Discussion

Although the sample as a whole made significant gains on receptive, productive and overall vocabulary gain, the effect sizes were small except for the overall gain of the long stay group (16-20 weeks). This finding provides tentative evidence in support of a *Duration Threshold Hypothesis* with regard to the impact of SA on gains in vocabulary size: that is, it may be the case that up to a certain point in time in the target language community the impact of being in that community on vocabulary gain is small, yet beyond that timeframe the impact becomes significantly larger. Segalowitz, Freed, Collentine, Lafford, Lazar and Díaz-Campos (2005) found large effect sizes for gains in speed and efficacy of L2 visual word recognition of SA learners of L2 Spanish whose length of stay in the target language community was 13 weeks; only slightly shorter than the long stay group of the present study. More evidence that may support a duration threshold with regard to SA and vocabulary gain was found by Ife, Vives Boix and Meara (2000), who found that length of stay was positively correlated with lexical organisation gains as measured by a word association task. Their long stay group stayed in the target language community for two academic semesters, which, if the learner travelled home for a fortnight or so between the semesters, is only slightly longer than 20 weeks (and therefore again closely matched to the long stay group of the present study).

Why might there be a duration threshold with regard to gains in vocabulary knowledge in a SA context? If, as posited by Swain's (1985) *Output Hypothesis*, production of lexis is the gateway to deep processing of it, and if, as posited by Hulstijn and Laufer's (2001) *Involvement Load Hypothesis*, vocabulary acquisition is mediated by the need for a lexical item, the search for that item and an evaluation of how to use the item, so it may be the case that the SA context is less facilitative in the short term of the conditions necessary for vocabulary size measurably to increase. In other words, it may take a longer stay in the target language community for opportunities to arise in which a learner might need, search for and use recently encountered lexical items.

The out-of-class language contact scenarios most highly identified with by the sample (shown in table 1) involved either receptive activity or interaction which was likely to be limited in terms of the lexis it generated (e.g. speaking to service personnel) and as such these scenarios may not offer the learner a wealth of opportunities for the use of recently encountered lexical items. It may be the case that as length of stay increases, so a learner further develops the confidence to try to manipulate the language contact scenarios they encounter in order to facilitate their use of recently encountered lexis. It may also be that as a learner's personal relationships with L2 speakers (e.g. their host family, their

fellow students) become deeper and more familiar over time, so the learner becomes more comfortable with producing language with those interlocutors and testing their fledgling L2 lexical hypotheses. Furthermore, as a learner's social network develops over time during study abroad, so they are likely to encounter increased opportunities to apply recently learned L2 lexis in a wider variety of discourse scenarios and with a wider range of interlocutors.

The long stay group outperformed the medium stay group on overall and receptive vocabulary gain. This finding supports those of Ife *et al.* (2000), Dwyer (2004), Félix-Brasdefer (2004), Sasaki (2009), and Llanes and Muñoz (2009), who all found a positive effect on linguistic gain of length of stay; that is to say, the longer a learner spends in the study abroad context, the higher their gains. In particular, Ife *et al.* (2000) found tentative evidence in support of the positive impact of greater length of stay on vocabulary development. DeKeyser (2014) argues for an interplay between length of stay and proficiency level; that is, that above a certain proficiency threshold acquisition slows and therefore if a learner is above that threshold, a longer length of stay is required for changes to be measurable. The medium stay participants in the present study had a higher mean proficiency level (as measured on arrival to the UK by the placement test at each participant's language institution) than the long stay group. Therefore, the finding that the long stay group outperformed the medium stay group on overall and receptive vocabulary gain may provide evidence in support of the linguistic threshold hypothesis as well as in support of the linguistic value of a longer length of stay.

The participants in Oxford identified more highly with scenarios involving watching television and reading subtitles than their counterparts in the capital (table 2). This may be because the Oxford participants were more likely to be living with native L2 speakers and, therefore, they were more likely to be living in a property that had a television set and TV license and in which English-language programmes would be chosen. The short stay participants identified more highly with out-of-class contact overall than the medium or long stay groups. This is surprising because one might assume that the short stay group would have experienced the least out-of-class contact. However, this finding may be attributable to a greater sense of urgency on the part of the short stay group to engage with the L2 based upon the shorter amount of time in which they had to do so.

Of the three out-of-class contact factors found, the scenarios most identified with by the sample as a whole were Interactive scenarios. The interactive scenarios all involved speaking and/or listening to spoken language outside of class, e.g. 'I listen to English songs outside of class' and 'I use English for short exchanges with the people I live with'. It should be noted that of the interactive scenarios, none involved speaking to interlocutors outside of the home or school network: this finding suggests that SA learners who live alone or with people who share their L1 are deprived of a key opportunity for out-of-class contact with the L2. In other words, if you do not live with L2 native speakers during SA, it is likely that you will miss out on the majority of the most salient out-of-class L2 contact scenarios.

The Interactive factor comprised six out-of-class contact scenarios of which four appeared in the top ten most identified with scenarios of both the present study and the only other study to have administered this version of the

LCP to a similar sample, that of Author (2015). These four scenarios included reading emails and webpages, reading timetables, posters and menus, watching films and videos, and getting information from strangers (e.g. directions / asking for the time). The shared features of these four scenarios reveal that learner-initiated obtainment of information from a variety of sources (online, 'real', interpersonal, auditory, visual etc.) pervades the kinds of out-of-class L2 contact that SA learners in England identify with.

The correlation analyses of out-of-class contact and vocabulary gain found no statistically significant relationship between the two. Similarly, the regression analyses indicated that the model did not significantly predict vocabulary gain. What might explain the lack of any relationship between out-of-class language contact and vocabulary gain in this study? It may be the case that the extent to which the participants identified with the out-of-class contact scenarios was insufficient to reveal a statistically significant relationship between these variables. That is to say, the sample may not have had sufficient out-of-class contact with English (or enough contact that was helpful in terms of acquiring new L2 lexis) for that contact to have significantly impacted on their vocabulary knowledge. As previously mentioned, the out-of-class contact scenarios most highly identified with by the sample involved either receptive activity (e.g. 'Read menus/timetables, posters etc.') or interaction which was likely to be limited in terms of the lexis it involves (e.g. 'Speak in English to service personnel') and as such these scenarios may not have offered the participants many opportunities to use recently encountered lexical items and therefore to increase the breadth of their vocabulary knowledge.

The regression analyses indicated that the only variable to predict vocabulary gain in this study was length of stay: belonging to the medium stay group (11-15 weeks) predicted lower receptive and overall vocabulary gain. The medium stay group had a significantly higher mean proficiency level than the long stay group. Therefore, an interplay between length of stay and proficiency may have negatively impacted upon the gains of the medium stay group: the threshold above which the proficiency level of this group stood may have inhibited measurable gains in receptive and overall vocabulary over the period of their stay in the target language community.

5. Conclusions

The sample in this study identified most highly with out-of-class language contact that involved self-initiated obtainment of information. These scenarios involved receptive activity or short interactional exchanges that were likely to be limited in terms of the lexis they might generate. This finding may explain the lack of a relationship identified between out-of-class contact and vocabulary gain in this study: in other words, the most identified with types of language contact may constrain SA learners in terms of opportunities for lexical input and use and therefore any possible link between out-of-class contact and lexical gain was obscured in this study. The pedagogical implication of this conclusion is clear: language institutions that host SA learners would do well to guide the out-of-class contact of their learners in order that their learners identify more highly with a wider range of L2 contact scenarios and thus have increased exposure to and opportunity to use the second language. Concordantly, Kinginger (2011:67)

highlights the important role the institution can play in promoting the engagement of SA learners in out-of-class L2 contact and encouraging them to take linguistic advantage of that contact. Indeed, there is evidence to show that curricular intervention results in increased engagement in out-of-class L2 contact: Dewey, Bown, Baker, Martinsen, Gold and Eggett (2014) compared the L2 contact of SA learners in Spain, Mexico, France, Russia, China and Egypt. Regression analysis indicated that studying on the Mexican and Egyptian programmes predicted higher out-of-class L2 contact ($p=.002$). Both of these programmes delivered organised opportunities for L2 contact beyond the classroom: the Mexican programme involved the participants in “regular volunteer activities” (2014:59) and the Egyptian learners were required to interact with native speakers outside of class for at least two hours daily and then discuss those interactions in regular interviews with programme leaders.

It may be that the out-of-class language contact the participants experienced in this study had an impact on dimensions of vocabulary knowledge other than breadth, such as depth and/or automaticity. Certainly, Fitzpatrick (2012) found that SA had a complex and non-linear impact on her participant’s acquisition of individual lexical items, and SA has been found to impact upon learners’ lexical organisation (Ife, Vives Boix & Meara, 2000). The use of a range of vocabulary measures to capture development across all of the dimensions of lexical knowledge is therefore a recommendation for future research.

Acknowledgements

This study was funded by the Economic and Social Research Council (ESRC).

REFERENCES

- Abdullah, K.I., Puteh, F., Azizan, A.R., Hamdan, N.N. & Saude, S. (2013). Validation of a controlled productive Vocabulary Levels Test below the 2000-word level. *System*, 41, 352-364.
- Author (2015)
- Bardovi-Harlig, K. (2002). A new starting point? Investigating formulaic use and input. *Studies in Second Language Acquisition*, 24, 189-198.
- Baró, A.L. & Serrano, R.S. (2011). Length of stay and study abroad: Language gains in two versus three months abroad. *Revista Española de Lingüística Aplicada*, 24, 95-110.
- Bartlett, M.S. (1954). A note on the multiplying factors for various chi square approximations. *Journal of the Royal Statistical Society*, 16 (Series B), 296-298.
- Beglar, D. (2000). Estimating vocabulary size. *JALT Testing & Evaluation SIG Newsletter*, 4, 2-3.
- Beglar, D. & Hunt, A. (1999). Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing*, 16, 131-162.
- Bundgaard-Nielsen, R.L., Best, C.T. & Tyler, M.D. (2011a). Vocabulary size matters: the assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, 32 (1), 51-67.
- Bundgaard-Nielsen, R.L., Best, C.T. & Tyler, M.D. (2011b). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, 33 (3), 433-461.
- Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition*, 26 (2), 227-248.
- Davidson, D.E. (2010). Language learning and study abroad. *Foreign Language Annals*, 43, 6-26.
- DeKeyser, R.M. (1986). *From learning to acquisition? Foreign language development in a U.S. classroom and during a semester abroad*. Unpublished doctoral thesis. Stanford University, Stanford, CA.
- DeKeyser, R.M. (1991). Foreign language development during a semester abroad. In B. Freed (Ed.), *Foreign language acquisition research and the classroom* (pp. 104-119). Lexington, Mass.: DC Heath and Co.
- DeKeyser, R.M. (2014). Methodological considerations about research on language development during study abroad". In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 313-326). Amsterdam: John Benjamins.
- Dewey, D.P. (2004). A comparison of reading development by learners of Japanese in intensive domestic immersion and study abroad contexts. *Studies in Second Language Acquisition*, 26, 303-327.
- Dewey, D.P. (2008). Japanese vocabulary acquisition and context. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 14, 127-148.
- Dewey, D.P., Belnap, R.K. & Hillstrom, R. (2013). Social network development, language use, and language acquisition during study abroad: Arabic language learners' perspectives. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 22, 84-110.
- Dewey, D.P., Bown, J., Baker, W., Martinsen, R.A., Gold, C. & Eggett, D. (2014). An exploratory study of language use during study abroad. *Language Learning*, 64 (1), 36-71.

- Dewey, D.P., Ring, S., Gardner, D. & Belnap, R. K. (2013). Social network formation and development during study abroad in the Middle East. *System*, 41, 269-282.
- Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition*, 26, 249-273.
- Dörnyei, Z. (2003). *Questionnaires in second language research*. Mahwah, NJ: Lawrence Erlbaum.
- Dwyer, M.M. (2004). Charting the impact of studying abroad. *International Education*, 4, 14-20.
- Eurolingua (2015). *Study abroad programs with Eurolingua* [Webpage]. Available at <http://www.eurolingua.com/> [Last accessed 27 April 2015].
- Félix-Brasdefer, J.C. (2004). Interlanguage refusals: Linguistic politeness and length of residence in the target community. *Language Learning*, 54 (4), 587-653.
- Fitzpatrick, T. (2012). Tracking the changes: Vocabulary acquisition in the study abroad context. *Language Learning Journal*, 40 (1), 81-98.
- Fitzpatrick, T. & Meara, P. (2004). Exploring the validity of a test of productive vocabulary. *Vigo International Journal of Applied Linguistics*, 1, 55-73.
- Foster, P. (2009) Lexical Diversity and native-like selection: the bonus of studying abroad. In H. Daller, D. Malvern, P. Meara, J. Milton, B. Richards & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: the interface between theory and application* (pp. 91-106). London: Palgrave-Macmillan.
- Freed, B.F. (1995). *Second language acquisition in a study abroad context*. Amsterdam, The Netherlands: John Benjamins Co.
- Freed, B.F., Dewey, D., Segalowitz, N. & Halter, R. (2004). The Language Contact Profile. *Studies in Second Language Acquisition*, 26 (2), 349-356.
- Freed, B.F., Segalowitz, N. & Dewey D.P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad and intensive domestic immersion programs. *Studies in Second Language Acquisition*, 26, 275-301.
- Freed, B.F., So, S. & Lazar, N. (2003). Language Learning Abroad: How Do Gains in Written Fluency Compare with Gains in Oral Fluency in French as a Second Language? *ADFL Bulletin*, 34 (3), 34-40.
- Hernández, T.A. (2010). The relationship among motivation, interaction, and the development of second language oral proficiency in a study-abroad context. *Modern Language Journal*, 94, 600-617.
- Højen, A. D. (2003). *Second-language speech perception and production in adult learners before and after short-term immersion*. Unpublished doctoral dissertation. University of Aarhus, Denmark.
- Hsueh-chao, M.H. & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13 (1), 403-430.
- Hulstijn, J.H., & Laufer, B. (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language Learning*, 51, 539-558.
- Ife, A., Vives Boix, G. & Meara, P. (2000). The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics*, 4, 55-84.
- Kinginger, C. (2008). Language learning in study abroad: Case studies of Americans in France. *Modern Language Journal*, 92, Monograph.
- Kinginger, C. (2011). Enhancing language learning in study abroad. *Annual Review of Applied Linguistics*, 31, 58-73.

- Koizumi, R. & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research*, 4, 900-913.
- Lafford, B. (2004). The effect of context of learning on the use of communication strategies by learners of Spanish as a second language. *Studies in Second Language Acquisition*, 26 (2), 201-26.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Bejoint & P. Arnaud (Eds.), *Vocabulary and Applied Linguistics* (pp. 126-132). Basingstoke and London: Macmillan.
- Laufer, B. & Nation, I.S.P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16, 307-322.
- Laufer, B. & Nation, I.S.P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16, 33-51.
- Laufer, B. & Paribakht, T.S. (1998). Relationship between passive and active vocabularies: effects of language learning context. *Language Learning*, 48, 365-391.
- Laufer, B. & Ravenhorst-Kalovski, G.C. (2010). Lexical threshold revisited: lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22 (1), 15-30.
- LearnEnglishGuide (2015). *Why learn English abroad* [Webpage]? Available at <http://www.learnenglishguide.com/abroad/> [Last accessed 27 April 2015].
- Llanes, À. & Muñoz, C. (2009). A short stay abroad: does it make a difference? *System*, 37 (3), 353-365.
- Magnan, S.S. & Back, M. (2007). Social interaction and linguistic gain during study abroad. *Foreign Language Annals*, 40, 43-61.
- Martinsen, R.A. (2011). Predicting changes in cultural sensitivity among students of Spanish during short-term study abroad. *Hispania*, 94, 121-141.
- Meara, P.M. & Jones, G. (1990). *The Eurocentres Vocabulary Size Tests: 10K*. Zurich: Eurocentres.
- Milton, J. & Meara, P.M. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL Review of Applied Linguistics*, 107/108, 17-34.
- Mora, J.C. (2008). Learning context effects on the acquisition of a second language phonology. In C. Pérez-Vidal, M. Juan-Garau & A. Bel (Eds.), *A portrait of the young in the new multilingual Spain* (pp. 241-63). Clevedon, England: Multilingual Matters.
- Mora, J.C. and Valls-Ferrer, M. (2012). Oral fluency, accuracy and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, 46 (4), 610-641.
- Muñoz, C. & Llanes, À. (2014). Study Abroad and Changes in Degree of Foreign Accent in Children and Adults. *The Modern Language Journal*, 98, 432-449.
- Myles, F., Hooper, J. & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning*, 48 (3), 323-363.
- Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Paribakht, T.S. & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: an introspective study of lexical inferencing. *Studies in Second Language Acquisition*, 21, 195-224.
- Pellegrino, V. (1998). Student perspectives on language learning in a study abroad context. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 4, 91-120.

- Pellegrino Aveni, V. (2007). Speak for yourself: second language use and self-construction during study abroad. In S. Wilkinson (Ed.), *Insights from study abroad for language programs* (pp. 99-115). Sydney: Thomson Heinle.
- Pizziconi, B. (2013). Japanese vocabulary development in and beyond study abroad: the timing of the year abroad in a language degree curriculum. *Language Learning Journal*. DOI: 10.1080/09571736.2013.786119.
- Platzer, H. (2012). *A CF approach to Nation's Vocabulary Levels Test*. Poster presented at the 9th EALTA Conference, Innsbruck.
- 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. 271-291). Amsterdam: John Benjamins.
- Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in Language Assessment* (pp. 41-60). Mahwah, NJ: Erlbaum.
- Rivers, W.P. (1998). Is being there enough? The effects of homestay placements on language gain during study abroad. *Foreign Language Annals*, 31, 492-500.
- Sasaki, M. (2009). Changes in EFL students' writing over 3.5 years: A socio-cognitive account. In R. Manchón (Ed.), *Learning, teaching and researching writing* (pp. 49-76). Clevedon: Multilingual Matters.
- Schmitt, N., Schmitt, D. & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18, 55-88.
- Segalowitz, N. & Freed, B. (2004). Context, contact, and cognition in oral fluency acquisition. *Studies in Second Language Acquisition*, 26, 175-201.
- Segalowitz, N., Freed, B., Collentine, J., Lafford, B., Lazar N. & Díaz-Campos, M. (2005). A comparison of acquisition of Spanish as a second language in two different contexts of learning: Study Abroad and the regular academic classroom. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 10, 1-18.
- Serrano, R., Llanes, À. & Tragant, E. (2011). Analyzing the Effect of Context of Second Language Learning: Domestic Intensive and Semi-Intensive Courses vs. Study Abroad in Europe. *System*, 39, 133-143.
- Staehr, L.S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36 (2), 139-152.
- Staehr, L.S. (2009). Vocabulary Knowledge and Advanced Listening Comprehension in English as a Foreign Language. *Studies in Second Language Acquisition*, 31 (4), 577-607.
- Swain, M. (1995). Three Functions of Output in Second Language Learning. In G. Cook & B. Seidlehofer (Eds.), *Principle and Practice in Applied Linguistics: Studies in Honour of H.G. Widdowson*. Oxford: Oxford University Press.
- Tanaka, K. (2004). *Changes in Japanese students' beliefs about language learning and English language proficiency in a study-abroad context*. Unpublished doctoral thesis. University of Auckland.
- Twombly, S.B. (1995). Piropos and friendship: Gender and culture clash in study abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 1, 1-27.
- van Zeeland, H. & Schmitt, N. (2013). Lexical coverage and L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34 (4), 457-479.
- Watson, J.R., Siska, P. & Wolfel, R.L. (2013). Assessing gains in language proficiency, cross-cultural competence, and regional awareness during study abroad: A preliminary study. *Foreign Language Annals*, 46 (1), 62-79.

- Wilkinson, S. (1998). Study abroad from the participants' perspective: A challenge to common beliefs. *Foreign Language Annals*, 31 (1), 23-39.
- Woodman, K. (1998). A study of linguistic, perceptual and pedagogical change in a short-term intensive language program. Unpublished doctoral thesis. University of Victoria, British Columbia, Canada.
- Xing, P. & Fulcher, . (2007). Reliability assessment for two versions of the Vocabulary Levels Test. *System*, 35 (2), 182-191.
- Yager, K. (1998). Learning Spanish in Mexico: The Effect of Informal Contact and Student Attitudes on Language ain. *Hispania*, 81 (4), 898-913.